



Universidade de Brasília
Faculdade de Economia, Administração e Contabilidade
Departamento de Economia
Mestrado Profissional em Economia do Setor Público

Estrutura a termo da taxa de juros no Brasil: Projeções utilizando aprendizado de máquina

Pedro Ivo Ferreira de Souza Junior

Brasília

2021

Pedro Ivo Ferreira de Souza Junior

**Estrutura a termo da taxa de juros no Brasil:
Projeções utilizando aprendizado de máquina**

Dissertação apresentada ao Programa de
Mestrado Profissional em Economia do Setor
Público, para a obtenção do título de Mestre.

Universidade de Brasília
Faculdade de Economia, Administração e Contabilidade
Departamento de Economia

Orientador: Prof. Daniel Oliveira Cajueiro, PhD

Brasília
2021

Pedro Ivo Ferreira de Souza Junior

**Estrutura a termo da taxa de juros no Brasil:
Projeções utilizando aprendizado de máquina**

Dissertação apresentada ao Programa de
Mestrado Profissional em Economia do Setor
Público, para a obtenção do título de Mestre.

Trabalho aprovado. Brasília 08 de julho de 2021

Prof. Daniel Oliveira Cajueiro, PhD
Universidade de Brasília
Orientador

Prof. Marina Delmondes de Carvalho Rossi, PhD
Universidade de Brasília

Helano Borges Dias, PhD

Brasília

2021

dD278e de Souza Junior, Pedro Ivo Ferreira
Estrutura a termo da taxa de juros no Brasil: Projeções utilizando aprendizado de máquina / Pedro Ivo Ferreira de Souza Junior; orientador Daniel Oliveira Cajueiro. -- Brasília, 2021
62 p.

Dissertação (Mestrado – Mestrado em Economia do Setor Público) -- Universidade de Brasília, 2021.

1. Economia. 2. Estrutura a termo da taxas de juros. 3. Aprendizado de máquina. I. Cajueiro, Daniel Oliveira, orient. II. Título.

Às minhas amadas esposa e filha: Raquel e Ana Carolina

AGRADECIMENTOS

Agradeço a Deus, primeiramente, pela vida e por todas as oportunidades que me concedeu.

À minha esposa, Raquel, e minha filha, Ana Carolina, pela compreensão e apoio constantes.

Aos meus pais, Pedro Ivo e Rose Mary, e irmãos, Anne Rose e Rafael, pelo incentivo e apoio incondicional em todos os momentos.

Ao meu orientador, Prof. Dr. Daniel Oliveira Cajueiro, pelas valiosas sugestões e pela orientação eficaz na execução deste trabalho.

À Prof. Dra. Marina Delmondes de Carvalho Rossi e ao Dr. Helano Borges Dias por terem aceitado o convite para compor a banca examinadora.

À Secretaria do Tesouro Nacional, em especial à Viviane Aparecida da Silva Varga, pelo apoio ao desenvolvimento deste trabalho.

RESUMO

Este trabalho tem como objetivo principal avaliar modelos de aprendizado de máquina, com o propósito de previsão de séries temporais, para projeção da Estrutura a Termo da Taxa de Juros – ETTJ brasileira. Os modelos tratados neste trabalho são: K Vizinhos mais Próximos, *Least Absolute Shrinkage and Selection Operator* – LASSO, Floresta Aleatória, *Extreme Gradient Boosting* e Máquinas de Vetor Suporte. No treinamento dos modelos foram utilizadas, primeiramente, as séries históricas das taxas de dezesseis vértices da ETTJ, e, adicionalmente, em um segundo momento, foram incluídas variáveis macroeconômicas aos dados de treinamento. O objetivo era avaliar se os desempenhos desses modelos seriam superiores ao do modelo Passeio Aleatório – PA, usado na literatura como referência para avaliação de desempenho de modelos de projeções de séries temporais. Os resultados mostraram que não é possível determinar um modelo, para os dados brasileiros, que consistentemente produza menor erro de previsão que o Passeio Aleatório em todos os horizontes de projeção. Por outro lado, quanto mais curto o prazo de projeção, mais os modelos analisados se tornam superiores ao Passeio Aleatório. As análises também indicaram que a inclusão de variáveis macroeconômicas para o treinamento não gerou melhores resultados nas projeções de forma consistente, apesar da existência de correlação entre essas variáveis e a dinâmica das taxas de juros da ETTJ.

Palavras-chave: Estrutura a Termo da Taxa de Juros. Modelos de Previsão. Aprendizado de Máquina.

Classificação JEL: C53, E43, E47, G17.

ABSTRACT

The main objective of this work is to evaluate machine learning models, with the purpose of forecasting time-series, for the forecasting of the Brazilian yield curve. The models covered in this work are K Nearest Neighbors, Least Absolute Shrinkage and Selection Operator – LASSO, Random Forest, Extreme Gradient Boosting and Support Vector Machines. In the model training, first, the historical time-series of the yield curve were used, for sixteen vertices, and, additionally, in a second moment, macroeconomic variables were included in the training data. The objective was to evaluate whether the performance of these models would be superior to that of the Random Walk model, used in the literature as a reference for evaluating the time-series performance of forecasting models. The results showed that it is not possible to determine a model, for the Brazilian data, that consistently produces less forecast error than the Random Walk in all projection horizons. On the other hand, the shorter the projection period, the more the analyzed models become superior to Random Walk. The results also indicated that the inclusion of macroeconomic variables for training did not consistently generate better results in the projections, despite the existence of a correlation between these variables and the dynamics of the yield curve.

Key-word: Yield Curve. Forecasting Models. Machine Learning.

JEL Classification: C53, E43, E47, G17.

LISTA DE FIGURAS

Figura 1 – Evolução da Curva de Juros	37
Figura 2 – Evolução das Taxas	37
Figura 3 – Evolução do Dólar e do CDS Brasil de 5 anos	39
Figura 4 – Evolução do IPCA e da Expectativa de IPCA 12 meses a frente	40
Figura 5 – Evolução da Taxa SELIC.....	40

TABELAS

Tabela 1 – Estatísticas Descritivas.....	38
Tabela 2 – Resultados Curvas de Juros, horizonte de 1 mês	50
Tabela 3 – Resultados Curvas de Juros, horizonte de 3 meses	51
Tabela 4 – Resultados Curvas de Juros, horizonte de 6 meses	52
Tabela 5 – Resultados Curvas de Juros, horizonte de 12 meses	52
Tabela 6 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 1 mês	53
Tabela 7 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 3 meses	54
Tabela 8 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 6 meses	55
Tabela 9 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 12 meses	55
Tabela 10 – TReREQM entre os modelos com diferentes covariadas	56
Tabela 11 – Teste de DM – Covariadas: Curvas de Juros	57
Tabela 12 – Teste de DM – Covariadas: Curvas de Juros e Variáveis Macroeconômicas	57

LISTA DE SIGLAS E ABREVIATURAS

ETTJ	Estrutura a Termo da Taxas de Juros
ML	Aprendizado de Máquina
PA	Passeio Aleatório
RF	Floresta Aleatória
LASSO	“Least Absolute Shrinkage and Selection Operator”
XGBoost	“Extreme Gradient Boosting”
kNN	K Vizinhos mais Próximos
SVM	Máquinas de Vetores Suporte
STN	Secretaria do Tesouro Nacional
DPF	Dívida Pública Federal
VAR	Vetor Auto Regressivo
DNS	“Dynamic Nelson-Siegel”
IA	Inteligência Artificial
ANBIMA	Associação Brasileira das Entidades dos Mercados Financeiros e de Capitais
LTN	Letras do Tesouro Nacional
NTN-F	Notas do Tesouro Nacional – Série F
IPCA	Índice Geral de Preços ao Consumidor
CDS	“Credit Default Swap”
PIB	Produto Interno Bruto
REQM	Raiz do Erro Quadrático Médio
ReREQM	Raiz do Erro Quadrático Médio Relativo
TREQM	Traço da Raiz do Erro Quadrático Médio
DM	Diebold-Mariano

SUMÁRIO

1. Introdução	13
2. Revisão Teórica Geral	16
2.1. Modelos de Estrutura a Termo de Taxas de Juros	16
3. Aprendizado de Máquina	20
3.1. Validação Cruzada	22
3.2. Otimização de Hiperparâmetros	23
4. Modelos de Aprendizado de Máquina	27
4.1. Passeio Aleatório	27
4.2. K Vizinhos mais Próximos	27
4.3. LASSO.....	28
4.4. Floresta Aleatória	30
4.5. “Extreme Gradient Boosting”	31
4.6. Máquinas de Vetores Suporte.....	33
5. Descrição dos Dados	36
5.1. Curva de Juros e Variáveis Macroeconômicas	36
5.2. Evolução da Economia e da Curva de Juros de 2007 a 2020.....	40
6. Resultados	47
6.1. Procedimentos de Previsão	47
6.2. Procedimentos de Avaliação dos Resultados.....	47
6.3. Resultados dos Modelos de Aprendizado de Máquina	49
7. Conclusão	58
8. Referências Bibliográficas	59

1. Introdução

A estrutura a termo da taxa de juros – ETTJ é a relação entre o retorno e o prazo de vencimento de instrumentos financeiros de renda fixa e fornece uma medida dos ganhos que um investidor pode ter em função de diferentes prazos de investimento. A ETTJ é conhecida como Curva de Juros quando apresenta as relações entre taxas e prazos em uma determinada data.

A modelagem da Curva de Juros e de sua dinâmica ao longo do tempo é importante para diversos atores, tais como: o mercado financeiro, famílias, firmas, economistas, gestores de passivos e de ativos. Esta modelagem é utilizada na definição dos preços de instrumentos de renda fixa e para a projeção de taxas futuras da curva de juros. A previsão da Curva de Juros é relevante para a economia porque a ETTJ contém as expectativas dos agentes de mercado com relação à trajetória futura da taxa de juros de curto prazo, definida pelo Banco Central para fins de condução da política monetária. Já do ponto de vista do emissor de títulos públicos, a curva de juros identifica o custo da dívida pública, tanto no curto quanto no longo prazo. Este custo é variável fundamental no processo de planejamento estratégico da gestão da dívida pública, para o cálculo do resultado fiscal e para a sustentabilidade do país, sendo um fator essencial para a tomada de decisão na condução da política fiscal.

A ETTJ também é importante para outros atores além do governo. Para as Instituições Financeiras, ela apresenta os fatores de desconto para precificação de ativos. Para a gestão de carteiras e entidades de previdência, a ETTJ é utilizada como parâmetro para a gestão de ativos e passivos, construção de *hedge* e para análises de exposição ao risco. No caso das famílias, a ETTJ afeta a decisão de poupar e consumir e, para as firmas, afeta na decisão de investir, uma vez que ela apresenta o custo de oportunidade de investimentos na economia real.

Em particular, a gestão da dívida no Brasil, feita na Secretaria do Tesouro Nacional – STN, utiliza projeções de curvas de juros no processo de planejamento estratégico da Dívida Pública Federal – DPF. As curvas de juros são utilizadas na precificação dos títulos presentes na estratégia de emissões, tanto para estratégias de curto quanto de longo prazo, permitindo uma análise da relação entre custo e risco de diversos instrumentos financeiros e de diferentes estratégias de financiamento. Outra aplicação da modelagem de curva de juros é na definição de curvas de juros de longo prazo, que são utilizadas no modelo de composição ótima de longo prazo da DPF. Projeções de curvas de juros também são importantes porque mostram

como o custo da dívida pode se comportar ao longo do tempo e ajudam na definição do custo de oportunidade na condução de políticas públicas.

Ao longo do tempo, foram propostos diversos modelos para modelar e prever a dinâmica da ETTJ. Estes modelos podem ser agrupados em três áreas: modelos baseados em não arbitragem, modelos de fatores e modelos de regressão da taxa *forward*. Contudo, a implementação dos modelos de previsão da ETTJ no Brasil é recente e não apresenta um modelo superior aos demais. Até 2002, as diversas crises econômicas pelas quais o Brasil passou, geraram volatilidade ao mercado de títulos públicos brasileiros, altas taxas de juros e prazos curtos de vencimentos, restringindo a capacidade de realização de trabalhos empíricos para a modelagem da dinâmica da ETTJ brasileira e sua projeção.

Desta forma, dada a sua importância, a introdução de novas tecnologias que permitam modelar, entender a dinâmica da curva de juros e projetá-la se tornam cada vez mais relevantes. Particularmente com relação a este último ponto, o aprendizado de máquina parece ser um instrumento apropriado para a projeção de curvas de juros, uma vez que ela modela relações não lineares e complexas, tais como a relação entre prazos de maturidade dos títulos e taxas de juros, e mesmo interações dessas taxas com variáveis macroeconômicas. Neste contexto, este trabalho tem o objetivo de utilizar modelos de aprendizado de máquina, com o propósito de previsão de séries temporais, para a projeção da ETTJ. Os modelos tratados neste trabalho são: K Vizinhos mais Próximos, *Least Absolute Shrinkage and Selection Operator* – LASSO, Floresta Aleatória, *Gradient Boosting Trees*, e Máquinas de Vetor Suporte. O objetivo principal do trabalho é verificar se, na previsão de Curvas de Juros, o desempenho do aprendizado de máquina é superior ao modelo Passeio Aleatório – PA, usado na literatura como referência para avaliação de desempenho de modelos de projeções de séries temporais.

Outro ponto a ser analisado neste trabalho, além do desempenho dos modelos de aprendizado de máquina para a projeção da ETTJ, é se a inclusão de algumas das principais variáveis macroeconômicas do país para treinamento desses modelos aumenta o desempenho dos mesmos. Principalmente após o processo de estabilização da economia no Brasil, permitiu-se a construção de séries temporais sem quebras estruturais. Após 2002, com a estabilização da economia, ocorreu um processo de alongamento dos títulos públicos no Brasil permitindo análises dessa natureza, onde o comportamento de variáveis macroeconômicas explica a variação das taxas da ETTJ ao longo do tempo.

Este trabalho, além da introdução, apresenta os seguintes capítulos:

- O segundo capítulo apresenta uma revisão bibliográfica dos modelos de previsão de curva de juros.
- O terceiro capítulo trata de conceitos relativos ao aprendizado de máquina.
- O quarto capítulo apresenta os modelos de aprendizado de máquina utilizados para as previsões, usando os dados brasileiros.
- O quinto capítulo descreve os dados utilizados para representar a curva de juros brasileira e expõe os eventos econômicos responsáveis pela dinâmica da curva de juros no Brasil.
- O sexto capítulo avalia o desempenho fora da amostra dos modelos de aprendizado de máquina.
- O sétimo capítulo apresenta as conclusões do trabalho, suas limitações e propõe questões para futuras pesquisas.

2. Revisão Teórica Geral

2.1. Modelos de Estrutura a Termo de Taxas de Juros

A explicação mais antiga e testada sobre a dinâmica das taxas de juros da economia é a Hipótese das Expectativas da ETTJ, formulada por Fischer (1896), que estabelece, em linhas gerais, que a taxa de juros de longo prazo é formada como uma média das taxas de juros de curto prazo esperadas para o futuro mais um prêmio de risco invariante no tempo. Assim, a diferença entre as taxas de curto e de longo prazo ETTJ refletiria as expectativas do mercado sobre as taxas de juros de curto prazo futuras, ou seja, um aumento da diferença entre a taxa longa e a taxa curta refletiria a expectativa de futuros aumentos nas taxas de curto prazo. Entretanto, estudos empíricos¹ para os dados dos Estados Unidos rejeitaram a Hipótese das Expectativas, pois se constatou que os prêmios de risco oscilam ao longo do tempo.

Posteriormente, foram desenvolvidos os modelos de não arbitragem, os quais incluem os modelos de equilíbrio. Nos modelos de equilíbrio, um conjunto de variáveis de estado descrevem a Curva de Juros, com essas variáveis sendo estimadas por meio de métodos estatísticos. Nestes modelos, também conhecidos como endógenos, a curva de juros em dado tempo t é derivada analiticamente a partir dos parâmetros definidos, contudo, não há garantias de que ela será idêntica à curva de mercado efetivamente observada.

Em seu artigo, Vasicek (1977) explicita as bases dos modelos de equilíbrio de um fator, descrevendo o processo para a taxa instantânea de juros a partir de uma equação diferencial estocástica. A fragilidade do modelo proposto por Vasicek decorre do fato de que o modelo pode gerar taxas reais de curto prazo negativas. Nesse sentido, Cox *et al.* (1985) propõem um procedimento alternativo, em que o processo para a taxa de juros de curto prazo é descrito por outra equação, onde um de seus componentes impede que as taxas de curto prazo se tornem negativas. Estes modelos foram seguidos por Brennan e Schwartz (1979), que assumiram que toda a ETTJ, em um determinado momento no tempo, pode ser expressa pelas taxas mais longa e mais curta dos títulos, e por Longstaff e Schartz (1992), que desenvolveram um modelo de equilíbrio de dois fatores, quais sejam: a taxa de juros de curto prazo e a volatilidade da taxa de juros de curto prazo.

Posteriormente, foram desenvolvidos os modelos livres de arbitragem. Nestes modelos, os parâmetros são calibrados de modo que o preço dos ativos nos modelos coincida

¹ Della Corte *et al.*, (2006)

com o seu preço de mercado, não permitindo, com isso, oportunidades de arbitragem. Entre estes modelos destacam-se: Ho e Lee (1986) e Hull e White (1990), com modelos de um único fator, e Black *et al.* (1990) e Heath *et al.* (1992), com modelos multifatoriais.

Dufee (2002), por sua vez, mostrou que estes modelos não apresentavam um bom resultado para a projeção das taxas dos títulos do mercado americano e que o modelo Passeio Aleatório – PA apresentava resultados superiores. A falha destes modelos era decorrente de uma de suas características: a compensação pelo risco é um múltiplo da variância do risco, o que limita a compensação que o investidor espera receber por assumir mais riscos. Outro fator que limitava a utilização destes modelos é que as curvas de juros apresentam uma grande variedade de formatos decorrente das combinações das taxas e prazos de maturidade, com uma dinâmica mais complexa que a apresentada pelos modelos.

Motivados pela complexidade das formas apresentadas pela ETTJ ao longo do tempo, Nelson e Siegel (1987) apresentaram um modelo com componentes exponenciais parametricamente estimados que tinha a capacidade de representar os diferentes formatos associados as curvas de juros quais sejam: crescente, decrescente, *humped* (corcova) e *inverted humped* (corcova invertida). Estes componentes são chamados de fatores latentes da curva de juros. O modelo apresenta três fatores que são associados ao nível, à inclinação e à curvatura das ETTJ, permitindo gerar formas complexas de curvas de juros. As curvas de juros geradas pelo modelo conseguiram prever as taxas dos juros de longo prazo com uma correlação de 0,96, o que sugeria que o modelo conseguiu capturar importantes atributos da relação taxas e prazos de vencimento. Posteriormente, Svensson (1995) estendeu o modelo de Nelson e Siegel com a adição de um novo termo exponencial à curva de juros, contendo dois parâmetros adicionais, permitindo que se forme uma segunda corcova na forma da curva de juros.

Diebold e Li (2006) utilizaram a forma funcional proposta por Nelson e Siegel para ajustar a dinâmica da curva de juros norte-americana da maneira mais suave possível e, com isso, extrair benefícios em termos de previsão fora da amostra. Os autores adaptaram o modelo anterior para um modelo dinâmico de três fatores chamado DNS (*Dynamic Nelson-Siegel*, em inglês). Este modelo tornou-se referência graças aos resultados obtidos para os dados americanos e por ter um desempenho superior ao do Passeio Aleatório.

Após Diebold e Li (2006) vários artigos foram apresentados com variações do modelo de Nelson-Siegel. Pooter (2007) apresentou um modelo de quatro fatores, assim como Svensson (1995). Koopman *et al.* (2010) apresentaram parâmetros que variam no tempo e

Diebold *et al.* (2008) trabalhou com dados de vários países, modelando parâmetros globais e específicos por país.

Outra linha de estudo inclui fatores macroeconômicos em conjunto com os fatores latentes para a projeção da ETTJ. Ang e Piazzesi (2003) trabalharam com um modelo de três fatores latentes em conjunto com fatores de inflação e crescimento econômico, estimados por um modelo de Vetor Auto Regressivo – VAR. O trabalho investigou como as variáveis macroeconômicas afetavam o preço dos títulos e a dinâmica da ETTJ. Os fatores macroeconômicos explicavam o movimento nos curto e médio prazos da curva enquanto os fatores não observáveis explicavam a maior parte do movimento na parte longa da curva de juros. Monch (2008) apresentou um modelo que usa a taxa de juros de curto prazo e várias variáveis macroeconômicas como fatores explicativos, conseguindo resultados significativos para prazos curtos e médios da curva de juros. Já Ludvigson e Ng (2009) usaram um modelo dinâmico de fatores e identificaram que o fator de inflação é um importante elemento para a previsão do prêmio de risco nos títulos do governo americano. Tanto Diebold *at al* (2006) quanto Pooter *et al.* (2010) também incluíram variáveis macroeconômicas em seus modelos e conseguiram resultados positivos com esta modelagem.

Uma abordagem diferente da apresentada por Nelson-Siegel foi implementada por Bowsher e Meeks (2008). Os autores apresentaram um modelo de cinco fatores, chamado de FSN-ECM, onde *splines* são usadas para modelar a ETTJ. O modelo apresenta flexibilidade, mas a interpretação dos fatores, de forma similar à apresentada por Nelson-Siegel, se perde.

Para o caso brasileiro, Guillen e Tabak (2009) testaram a Hipótese de Expectativa para a ETTJ e a rejeitaram ao identificar prêmios de risco que variavam ao longo do tempo.

Já Cajueiro *et al.* (2009) utilizou os modelos DNS e FSN-ECM (respectivamente apresentados por Diebold e Li (2006) e Bowsher e Meeks (2008)) em comparação com o modelo de Passeio Aleatório. Os resultados mostraram que o modelo FSN-ECM apresentavam resultados superiores para o curto prazo enquanto o modelo DNS produziu melhores estimativas para prazos mais longos.

Por sua vez, Leite *et al.* (2009) implementou um modelo com variáveis macroeconômicas para previsão da ETTJ, encontrando resultados superiores ao PA e aos modelos baseados em Nelson-Siegel. O modelo utilizou a expectativa de inflação e o prêmio de risco das taxas *forwards*.

Por fim, Araújo (2011), buscou identificar a existência de técnicas de combinação de modelos que possibilitassem a obtenção de projeções da ETTJ superiores às dos modelos

individuais para os dados brasileiros. Os resultados alcançados levaram à conclusão que não era possível identificar um modelo individual que consistentemente produzisse menor erro de previsão para o período fora da amostra, similar ao encontrado por Cajueiro *et al.* (2009). Adicionalmente, em cada horizonte de previsão considerado, foi possível encontrar pelo menos um esquema de combinação que tenha apresentado desempenho igual ou superior ao melhor modelo individual e, quanto mais longo o prazo considerado, maior a contribuição das combinações.

3. Aprendizado de Máquina

A Inteligência Artificial – IA é um campo da ciência cujo propósito é estudar, desenvolver e empregar máquinas para realizarem atividades humanas de maneira autônoma. Ela está relacionada à Robótica, ao Aprendizado de Máquina, ao reconhecimento de voz e imagens, dentre outras tecnologias.

Por sua vez, o Aprendizado de Máquina – ML (*Machine Learning* – ML, em inglês) é uma vertente específica da IA que treina as máquinas, através de algoritmos, para aprender com dados. O Aprendizado de Máquina envolve métodos de avaliação de dados que automatiza o desenvolvimento e identificação de padrões analíticos.

Ele tem como base a concepção de que sistemas tecnológicos podem aprender usando dados, de modo a descobrir padrões, tomar decisões e se aperfeiçoar com pouca interferência humana. Desta forma, pode, inclusive, melhorar a realização de uma atividade ao longo do tempo.

As tarefas de Aprendizado de Máquina são tipicamente classificadas em categorias de acordo com a natureza do processo de aprendizado, quais sejam: aprendizado supervisionado, não supervisionado, semi-supervisionado e por reforço.

No aprendizado supervisionado, os modelos são treinados por meio de exemplos com entrada de dados para os quais a saída desejada é conhecida. O algoritmo de aprendizado supervisionado recebe um conjunto de entradas e suas respectivas saídas e aprende ao comparar a saída do modelo com as saídas corretas, avaliando os erros entre esses dois conjuntos de resultados. O algoritmo, então, modifica o modelo de acordo com os dados. O aprendizado supervisionado utiliza padrões para prever os valores de rótulos em dados adicionais não rotulados. O aprendizado supervisionado é geralmente empregado em aplicações nas quais dados históricos preveem eventos futuros prováveis.

Já o aprendizado não supervisionado é utilizado em dados que não possuem as saídas correspondentes aos dados de entrada. A resposta não é informada ao sistema. O algoritmo deve descobrir os padrões que estão nos dados. O objetivo é explorar os dados e encontrar alguma estrutura interna. Técnicas populares incluem mapas auto-organizáveis, mapeamento por proximidade, agrupamento *k-means* e decomposição em valores singulares. Esses algoritmos também são utilizados para segmentar tópicos de texto, recomendar itens e identificar pontos discrepantes nos dados.

O aprendizado semi-supervisionado é utilizado para as mesmas aplicações que as do aprendizado supervisionado. Mas esse manipula tanto dados rotulados quanto não-rotulados para treinamento – normalmente uma pequena quantidade de dados rotulados com uma grande quantidade de dados não-rotulados, uma vez que dados sem rótulos são mais baratos e demandam menos esforços para serem adquiridos. Esse tipo de aprendizado pode ser empregado com métodos como classificação, regressão e previsão. O aprendizado semi-supervisionado é útil quando o custo associado à rotulação é muito alto para possibilitar um processo de treinamento totalmente rotulado.

Por fim, no aprendizado por reforço, o algoritmo descobre através de testes do tipo 'tentativa e erro' quais ações rendem as maiores recompensas. Este tipo de aprendizado possui três componentes principais: o agente (aprendiz ou tomador de decisão), o ambiente (tudo com que o agente interage) e ações (o que o agente pode fazer). O objetivo é que o agente escolha ações que maximizem a recompensa esperada em um período de tempo determinado. O agente atingirá o objetivo mais rápido se seguir uma boa política. Desta forma, o foco do aprendizado por reforço é descobrir esta melhor política.

O Aprendizado de Máquina e a estatística são campos intimamente relacionados. A diferença fundamental entre inferência estatística e aprendizado de máquina está relacionada às quantidades de interesse que devem ser otimizadas em função da informação contida nas bases de dados e ao tipo de resultado produzido.

Na inferência estatística, os parâmetros estimados são as estatísticas que tentam descrever o valor de suas contrapartidas na população, através da informação de uma amostra. Neste sentido, os parâmetros são o resultado de um processo de estimação. Já no Aprendizado de Máquina, as quantidades de interesse são os hiperparâmetros, variáveis que controlam o próprio processo de treinamento. Portanto, os hiperparâmetros são escolhidos antes da aplicação do algoritmo aos dados. Sem essa escolha, o algoritmo não é capaz de produzir resultados a partir dos dados, ao contrário dos parâmetros que resultam automaticamente de uma especificação original.

Desta forma, a principal diferença entre essas duas áreas é com relação aos seus propósitos. No caso da inferência estatística, o interesse está na interpretação dos parâmetros do modelo, permitindo, com isso, a escolha de Políticas Públicas. Por sua vez, o propósito do Aprendizado de Máquina é fazer projeções.

Escolher a especificação de um modelo cujos parâmetros devem ser estimados equivale a escolher um conjunto inicial de hiperparâmetros, antes de avaliar os dados. Nos

dois casos, os resultados ainda dependem da aplicação do algoritmo ao conjunto de informações, e os resultados obtidos dependem da informação presente nos dados, mas de forma condicional às escolhas realizadas.

No caso de inferência, os produtos da estimação são os parâmetros estimados, e eventuais previsões fora da amostra são feitas a partir desses valores, contudo, as projeções não são o seu principal objetivo. Já no caso dos algoritmos de Aprendizado de Máquina e seus hiperparâmetros, os resultados são os valores previstos diretamente. Enquanto valores previstos são obtidos a partir de um segundo estágio no caso de inferência, agora são os efeitos parciais os potenciais resultados obtidos em um segundo passo dos modelos de Aprendizado de Máquina.

3.1. Validação Cruzada

A validação cruzada é um método de reamostragem e tem como objetivo avaliar a capacidade de generalização de um modelo. No caso da previsão de séries temporais, o desenvolvimento e o uso de um modelo preditivo envolvem dados nos quais há um relacionamento ordenado entre as observações. Adicionalmente, ao avaliar um modelo para previsão de série temporal, estamos interessados no desempenho do modelo em dados que não foram usados para treiná-lo.

Esta avaliação pode ser feita dividindo os dados disponíveis em dois agrupamentos. O primeiro agrupamento, chamado de conjunto de treinamento, é utilizado para treinar o modelo e o segundo agrupamento, denominado conjunto de teste, é utilizado para avaliação das previsões. A avaliação dessas previsões fornece uma ideia de como o modelo funcionará com novas amostras.

O método de validação cruzada denominado *k-fold* consiste em dividir o conjunto total de dados em k subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir daí, um subconjunto é utilizado para teste, já os $k - 1$ dados restantes são utilizados para estimação dos parâmetros, fazendo-se o cálculo da acurácia do modelo. Este processo é realizado k vezes alternando de forma circular o subconjunto de teste.

Contudo, este método não pode ser usado diretamente em séries temporais, uma vez que ele assume que não há relação entre as observações, ou seja, cada observação é independente uma da outra. Ou seja, quando a dimensão temporal das observações importa,

não podemos dividi-las aleatoriamente em grupos. Em vez disso, devemos dividir os dados e manter a ordem temporal em que os valores foram observados.

Existem duas abordagens para a validação cruzada para séries temporais: a da janela deslizante e a da janela crescente.

No método da janela deslizante, o treinamento é feito em n dados e a validação é feita nos próximos m dados. Para o próximo período, essa janela de $n + m$, que tem tamanho fixo, é deslocada um período à frente e o treinamento e a validação são repetidos. Desta forma, a ordem entre as observações são preservadas.

Já no método de janela crescente, o treinamento é feito nos últimos n períodos e a validação ocorre nos próximos m períodos. Neste caso, a janela de $n + m$ períodos, que apresenta tamanho crescente, também é deslocada à frente, com os dados de treinamento aumentando sempre um período, decorrente da inclusão de um dado na janela de treinamento.

No caso de séries temporais, além da projeção de um período à frente, também é relevante efetuar projeções n períodos à frente. Esse caso também pode ser implementado usando os dois métodos descritos nos parágrafos anteriores.

3.2. Otimização de Hiperparâmetros

No aprendizado de máquina, a otimização de hiperparâmetros (ou ajuste de hiperparâmetros) é o problema de se escolher um conjunto de hiperparâmetros ideais para um determinado algoritmo de aprendizado. Conforme definido anteriormente, um hiperparâmetro é um parâmetro cujo valor é usado para controlar o processo de aprendizagem. Por outro lado, os valores de outros itens, como, por exemplo, os pesos dos nós, são ajustados através do algoritmo de aprendizagem.

Conforme Bergstra e Bengio (2012), o objetivo de um algoritmo de aprendizado \mathcal{A} , é encontrar uma função f , que minimize uma perda esperada, $L(x: f)$, sobre as amostras x de uma distribuição natural \mathcal{G}_x . Um determinado tipo de modelo de aprendizado de máquina pode exigir a determinação de diferentes valores de hiperparâmetros λ . O modelo é definido após a escolha dos λ , sendo denotado como \mathcal{A}_λ , e, também, pela determinação de $f = \mathcal{A}_\lambda(\mathcal{X}^{(treino)})$, para um conjunto de treinamento $\mathcal{X}^{(treino)}$, de forma a generalizar diferentes padrões de dados. Essas variáveis precisam ser escolhidas para que o modelo possa resolver de maneira ótima o problema de aprendizado de máquina. A otimização encontra uma tupla de hiperparâmetros que produz um modelo ótimo, que, por sua vez, minimiza uma função de

perda, predefinida com dados independentes fornecidos, ou seja, a otimização escolhe λ de forma a minimizar o erro $\mathbb{E}_{x \sim \mathcal{G}_x}[L(x; \mathcal{A}_\lambda(\mathcal{X}^{(treino)}))]$. O processo de otimização consiste em encontrar a solução para o seguinte problema de minimização:

$$\lambda^* = \min_{\lambda \in \Lambda} \mathbb{E}_{x \sim \mathcal{G}_x}[L(x; \mathcal{A}_\lambda(\mathcal{X}^{(treino)}))] \quad (1)$$

A função objetivo pega uma tupla de hiperparâmetros e retorna a perda associada. Como visto na seção anterior, a validação cruzada é frequentemente usada para estimar esse desempenho de generalização.

Na prática, o processo de otimização dos hiperparâmetros é obtido através das equações (2) a (4):

$$\lambda^* \approx \min_{\lambda \in \Lambda} \text{media}_{x \in \mathcal{X}^{(v\u00e1lido)}} L(x; \mathcal{A}_\lambda(\mathcal{X}^{(treino)})) \quad (2)$$

$$\equiv \min_{\lambda \in \Lambda} \Psi(\lambda) \quad (3)$$

$$\approx \min_{\lambda \in \{\lambda^1 \dots \lambda^S\}} \Psi(\lambda) \equiv \hat{\lambda} \quad (4)$$

A equação (3) apresenta o problema de otimização em termos de uma função resposta dos hiperparâmetros, Ψ , que ocorre através da minimização de $\Psi(\lambda)$ dado $\lambda \in \Lambda$. Como normalmente pouco se sabe sobre a função resposta Ψ e sobre o espaço de procura Λ , a estratégia para se encontrar um λ ótimo é encontrar um conjunto (S) de possibilidades $\{\lambda^1 \dots \lambda^S\}$, avaliar $\Psi(\lambda)$ para cada ponto possível e retornar o λ^i que melhor atuaria como $\hat{\lambda}$. Este método é denominado *grid search*.

Ou seja, no *grid search*, os hiperparâmetros são definidos através de uma busca exaustiva entre todas as possíveis combinações de um subconjunto de valores definidos previamente, sendo, portanto, um processo computacionalmente intensivo, uma vez que o conjunto de valores cresce exponencialmente com o aumento na quantidade de hiperparâmetros. O algoritmo de *grid search* utiliza uma métrica de desempenho, normalmente medida por validação cruzada no conjunto de treinamento, para identificar o conjunto ótimo de hiperparâmetros.

Outro algoritmo usado na otimização de hiperparâmetros é o *random search*, que se difere do *grid search* por procurar, de forma iterativa, melhores resultados no espaço de busca, sendo que esta procura ocorre através da obtenção de uma nova amostra de valores calculada através da soma de um vetor de valores aleatórios relativos à posição inicial. O algoritmo recebe, ao invés de um conjunto de possíveis valores para os parâmetros, um

intervalo para os possíveis valores. Enquanto o *grid search* compara os resultados de todas as combinações possíveis, o *random search* parte de uma posição inicial neste intervalo e, se o novo conjunto de valores obtidos aleatoriamente para os hiperparâmetros apresentar melhor resultado, o algoritmo se desloca para essa nova posição. Bergstra e Bengio (2012) demonstram que o *random search* é, tanto empiricamente quanto teoricamente, mais eficiente que o *grid search*.

Por fim, o processo de definição dos hiperparâmetros também pode ser feito através de um algoritmo de otimização Bayesiana. A otimização Bayesiana é um método de otimização global para funções de caixa preta² ou que apresentam um alto custo de avaliação. A otimização Bayesiana constrói um modelo de probabilidade da função objetivo, chamado de modelo de superfície de resposta (ou função objetiva substituta³), que aproxima a verdadeira função objetivo, e a usa para selecionar hiperparâmetros que serão, em seguida, avaliados na verdadeira função objetivo, de forma a encontrar o máximo/mínimo global das amostras avaliadas.

Busca-se, portanto, otimizar uma função não linear $f(x)$ sob um conjunto compacto A :

$$\min_{x \in A} f(x) \quad (5)$$

A otimização Bayesiana é uma abordagem que utiliza o Teorema de Bayes para direcionar a busca, a fim de se encontrar o mínimo ou máximo de uma função objetivo. O Teorema de Bayes é utilizado para calcular a probabilidade condicional de um evento:

$$P(A|B) = P(B|A) \frac{P(A)}{P(B)} \quad (6)$$

Esse cálculo pode ser simplificado removendo-se o valor de normalização de $P(B)$ e descrevendo a probabilidade condicional como uma quantidade proporcional. Esta forma é útil quando não há interesse em calcular uma probabilidade condicional específica, mas sim em otimizar uma determinada quantidade:

² Caixa preta um sistema fechado de alta complexidade, no qual a sua estrutura interna é desconhecida ou não é levada em consideração em sua análise, que se limita, assim, a medidas das relações de entrada e saída.

³ Aproximação bayesiana da função objetivo cujas amostras são obtidas de forma eficiente

$$P(A|B) = P(B|A)P(A) \quad (7)$$

Na equação (7), a probabilidade condicional $P(A|B)$ é chamada de probabilidade posterior, a probabilidade marginal $P(A)$ é definida como a probabilidade anterior, e a probabilidade condicional reversa $P(B|A)$ é a verossimilhança. A equação 7 apresenta a solução para otimizar a função objetivo. Primeiro são criadas amostras dos hiperparâmetros (x_1, x_2, \dots, x_n) que são avaliadas usando a função objetivo $f(x_i)$. Essa função retorna o custo ou resultado para cada amostra x_i . As amostras e seus resultados são coletados sequencialmente e definem os dados, $D_{1:n} = \{x_{1:n}, f(x_{1:n})\}$, que serão usados para definir a probabilidade anterior, $P(f)$. A função de verossimilhança é definida como a probabilidade de observar os dados dada a função $P(D_{1:n}|f)$. Esta função de verossimilhança mudará à medida que mais observações forem coletadas. Desta forma, os resultados da probabilidade anterior são combinados com a função de verossimilhança para se obter a distribuição da probabilidade posterior:

$$P(f|D_{1:n}) = P(D_{1:n}|f)P(f) \quad (8)$$

A probabilidade posterior $(f|D_{1:n})$ é a função objetiva substituta, ou seja, é uma aproximação da função objetivo original.

Para obter os valores do hiperparâmetros de forma eficiente, a otimização Bayesiana utiliza a função substituta em uma função chamada de função de aquisição (ou função de seleção), que representa o critério que será utilizado para determinar o próximo valor do hiperparâmetro $(x_t + 1) \in A$. Esta técnica de otimização tem a propriedade de minimizar o número de avaliações da função objetivo. Além disso, ela apresenta um bom desempenho, mesmo em ambientes onde a função objetivo tem vários máximos locais.

4. Modelos de Aprendizado de Máquina

A primeira etapa deste trabalho reside na definição e estimação dos modelos de aprendizado de máquina para a projeção da curva de juros. Os modelos considerados divergem quanto à sua complexidade e quanto ao custo computacional para sua estimação. Os modelos considerados são: K Vizinhos mais Próximos, *Least Absolute Shrinkage and Selection Operator* - LASSO, Florestas Aleatórias, *Gradient Boosting Trees*, e Máquinas de Vetores Suporte, que terão seu desempenho comparado ao modelo Passeio Aleatório.

4.1. Passeio Aleatório

O modelo mais simples e que será o utilizado como referência para a avaliação do desempenho dos outros modelos é o Passeio Aleatório.

O modelo de passeio aleatório das taxas de juros para cada maturidade τ_i é dado por:

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t(\tau_i) \quad (9)$$

Onde $\varepsilon_t(\tau_i) \sim N(0, \sigma^2(\tau_i))$ é um ruído branco.

Já a previsão das taxas, h períodos à frente, é dada por:

$$\hat{y}_{t+h}(\tau_i) = y_t(\tau_i) \quad (10)$$

Logo, a previsão é igual ao valor mais recente observado.

4.2. K Vizinhos mais Próximos

Um dos modelos mais simples para abordar tanto problemas de classificação quanto de regressão é o k Vizinhos mais Próximos – kNN (*k-Nearest Neighbors* – kNN em inglês).

Esse método é do tipo não paramétrico e onde, também, não é feita qualquer suposição a respeito dos dados.

O kNN requer todos os padrões de treinamento $x_i \in R^{K \times 1}$, em conjunto com as respectivas respostas $y(i)$, $i = 0, \dots, N - 1$. Logo, para cada dado novo de entrada x' , a

saída do modelo depende das saídas associadas aos k padrões de treinamento que estão mais próximas à entrada x' . Pode-se, por exemplo, utilizar a média simples das respostas dos vizinhos mais próximos:

$$\hat{y}(x') = \frac{1}{k} \sum_{x(i) \in N_k(x')} y(i) \quad (11)$$

Onde $N_k(x')$ é a vizinhança de x' , formada pelos padrões de treinamento $x(i)$ que correspondem aos k vizinhos mais próximos.

Desta forma, o modelo envolve a definição de uma métrica de distância para determinar os vizinhos mais próximos e um valor para o hiperparâmetro k , que é o número de vizinhos que são levados em consideração no cálculo da saída.

Em função de suas características, o kNN é definido como um método de aprendizado competitivo, e pertence a uma classe de algoritmos chamados de preguiçosos, uma vez que ele não constrói um modelo até o momento em que a predição é necessária. Com isso, o método tem o benefício de incluir apenas os dados relevantes para a análise do novo padrão de entrada. A desvantagem é que os dados de treinamento devem ser armazenados e consultados para a identificação dos vizinhos mais próximos.

No caso de regressão, a saída do modelo para um novo dado de entrada x' pode ser escrita como:

$$\hat{y}(x') = \frac{\sum_{j=1}^k w_j y_j(x \in N_k(x'))}{\sum w_j} \quad (12)$$

Onde $w_j, j = 1, \dots, k$ representa o peso associado ao $j = \text{ésimo}$ vizinho de x' .

4.3. LASSO

O modelo LASSO é uma variação da regressão linear e é utilizado quando há um grande número de variáveis preditoras, uma vez que esse método executa tanto um processo de seleção de variáveis quanto de regularização, de forma a aumentar a acurácia das projeções e facilitar a interpretação do modelo estatístico.

No caso da regressão linear, os parâmetros β são estimados e as variáveis associadas às estimativas desses parâmetros que são diferentes de zero entrarão no modelo estatístico. Em certas condições, os estimadores de Mínimos Quadrados Ordinários (MQO) são não

viesados e com mínima variância, ou seja, são estimadores não viciados e com menor variabilidade.

Contudo, como o objetivo do aprendizado de máquina é a projeção de novas observações, pode-se permitir certo viés nas estimativas dos parâmetros, deixando de selecionar algumas variáveis, de forma a aumentar a acurácia do modelo.

A seleção de variáveis é o processo de escolha de um subconjunto de características que são relevantes (variáveis e preditores) em um modelo estatístico, partindo da premissa que os dados contêm características que são redundantes ou irrelevantes para a definição deste modelo e cuja remoção pode ser efetuada sem grande perda de informação. A seleção de variáveis permite a simplificação dos modelos, reduz o tempo de treinamento, ajuda a evitar a maldição da dimensionalidade e a reduzir o problema de sobreajustamento (*overfitting*⁴).

Por sua vez, a regularização é o processo de adicionar uma restrição, ou seja, um termo de penalidade, às estimativas dos parâmetros de MQO. Esta restrição impõe um custo à função de otimização ao se estimar a solução do problema a seguir:

$$\hat{\beta}^{restrito} = \min_{\beta} \sum_i^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \quad (13)$$

Sujeito a $g(\beta) < t$, em que $g(\beta)$ representa a função de penalidade.

A função de penalidade tem a função de manter as estimativas de β_j próximas à zero.

O processo de penalização é implementado através de Multiplicadores de Lagrange, aumentando a função objetivo, conforme equação abaixo:

$$\hat{\beta}^{restrito} = \min_{\beta} \sum_i^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda g(\beta) \quad (14)$$

Em que λ é um escalar entre zero e infinito. À medida que λ cresce, a variância do modelo é reduzida e se aumenta o vício.

Quando a penalização é a LASSO, temos a seguinte especificação:

$$\hat{\beta}^{Lasso} = \min_{\beta} \sum_i^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|^q \quad (15)$$

Onde $q = 1$.

⁴ *Overfitting* ocorre quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados.

Dependendo do valor de λ , algumas estimativas dos parâmetros são efetivamente zero, com isso, o penalizador Lasso também funciona como um selecionador de variáveis. Finalmente, no caso do modelo Lasso, o hiperparâmetro a ser otimizado é o λ .

4.4. Floresta Aleatória

Floresta Aleatória – RF (*Random Forest* – RF, em inglês) é um método de aprendizado de máquina, que utiliza a técnica de sistemas de classificadores múltiplos (*ensemble learning*), através da construção de múltiplas Árvores de Decisão, que também é um modelo de aprendizado de máquina. A ideia é que a construção de várias Árvores de Decisão, a partir de subconjuntos de características dos dados, gerem respostas que, combinadas, possam levar a uma melhor generalização e obtenham resultados superiores ao modelo que usa somente uma árvore.

Por sua vez, a Árvore de Decisão é um modelo de predição cuja estrutura é semelhante à de uma árvore e que representa uma tabela de decisões. Este modelo mapeia diferentes alternativas para a tomada de decisão, com seus respectivos resultados e probabilidades de ocorrência. Cada nó da árvore envolve um questionamento sobre um dos atributos presentes nos dados. Os resultados desses questionamentos dividem os dados entre os ramos da árvore, permitindo o agrupamento de dados com características semelhantes. Os nós finais de cada ramo são chamados de folhas e definem uma classe de dados, representando as decisões tomadas após a identificação dos atributos dos dados. Com isso, novos dados são classificados com os atributos presentes nos dados de treinamento de cada folha. No caso de regressão, é atribuída ao novo dado a média dos valores dos elementos das folhas.

A Árvore de Decisão, apesar de ser um algoritmo intuitivo e robusto, é susceptível a sobreajustamento, uma vez que, para gerar bons resultados, usualmente são necessárias árvores grandes e com muitos nós, ocorrendo, com isso, ajuste excessivo nos dados de treinamento e resultando em uma baixa capacidade preditiva em novos dados. Essa fragilidade é reduzida no modelo de Florestas Aleatórias com a utilização de técnicas de sistemas de classificadores múltiplos, limitando o tamanho das árvores e reduzindo-se, com isso, a possibilidade de sobreajustamento.

Uma das formas de se construir um *ensemble* de árvores é através da técnica de *bootstrap aggregating*, ou *bagging*. O *bootstrap* é um método de geração de amostras em que o número de objetos escolhidos é o mesmo que o número inicial de objetos. Mas esses objetos

são escolhidos com substituições. Em outras palavras, um conjunto de dados de tamanho N é amostrado (com reposição), gerando M “novos” conjuntos que, por sua vez, são usados para construir M árvores. As respostas dessas árvores são combinadas para gerar a saída do *ensemble*. A construção de um *ensemble* em Florestas Aleatórias envolve a criação de Árvores de Decisão a partir de subconjuntos de características dos dados. Em Ho (1995), discute-se que, se houver m atributos, haverá 2^m possíveis subconjuntos, gerando diversidade, o que geralmente leva a melhores modelos.

O algoritmo de treinamento da Floresta Aleatória é descrito em seguida. Dado um conjunto de treinamento $X = x_1, \dots, x_n$ com respostas $Y = y_1, \dots, y_n$, selecionados de forma repetida A vezes, seleciona-se uma amostra aleatória do conjunto de treinamento com repetição e, com essas amostras, são realizados os treinamentos das árvores. Para $a = 1, \dots, A$ efetua-se amostragem, com repetição, de n exemplos de treinamento $X_a Y_a$. Em seguida, é treinada uma árvore f_a em $X_a Y_a$. As previsões para novos dados são feitas calculando-se a média das projeções de todas as árvores de regressão individuais:

$$\hat{f} = \frac{1}{A} \sum_{a=1}^A f_a(x) \quad (16)$$

A técnica de *bagging* melhora o desempenho do modelo, reduzindo sua variância sem aumentar o seu viés, ou seja, a média das árvores não é sensível ao ruído em seu conjunto de treinamento, desde que as árvores não sejam correlacionadas.

Um algoritmo de Floresta Aleatória tem todos os hiperparâmetros de uma Árvore de Decisão e também todos os hiperparâmetros de um classificador de *bagging*, para controlar a combinação de árvores.

4.5. “Extreme Gradient Boosting”

O *Extreme Gradient Boosting* – XGBoost é uma implementação do modelo *Gradient Boosting* e que apresenta melhorias em relação a outros algoritmos de reforço de gradiente. Já o *Gradient Boosting* é um modelo construído na forma de um *ensemble* de modelos de previsão fracos, geralmente Árvores de Decisão.

Assim como no *bagging*, os métodos de *boosting* treinam vários modelos mais simples com a finalidade de produzir um modelo final mais robusto, ou seja, converte uma

aprendizagem fraca em uma aprendizagem forte. No entanto, nos algoritmos de *boosting*, os modelos não são mais treinados de forma independente entre si, mas de maneira sequencial, a partir de um ajuste dos modelos treinados previamente. O algoritmo treina novos modelos diretamente no erro dos modelos treinados anteriormente. Ou seja, os novos modelos tentam prever o erro dos modelos anteriores em vez de prever independentemente o resultado. Dessa forma, a projeção final é obtida somando a predição de todos os modelos de previsão fracos.

Desta forma, o *Gradient Boosting* é um algoritmo de descida de gradiente funcional, ou seja, é um método que otimiza uma função de custo sobre um espaço de funções, escolhendo iterativamente uma função (hipótese fraca) que aponta na direção oposta do gradiente. Se a função é convexa, o procedimento converge para o ponto de mínimo da função.

O algoritmo do *Gradient Boosting* é apresentado a seguir na forma de um pseudocódigo, e foi extraído do livro de Hastie, Tibshirani e Friedman (2001):

$$1) \quad f_0(x) = \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$$

Inicializa-se a primeira árvore (iteração), onde $L(f)$ é função de custo.

$$2) \quad \text{Para } m = 1 \text{ até } M:$$

Onde M é o total de árvores (ou iterações).

$$a) \quad \text{Para } i = 1, 2, \dots, N \text{ calcule:}$$

$$r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}$$

Ou seja, para cada observação i , calcula-se o resíduo dado pelo negativo do gradiente da função de custo avaliada em f_{m-1} .

b) Treine uma árvore de regressão sobre a variável resposta de resíduo r_{im} , resultando nas regiões R_{jm} , $j = 1, 2, \dots, J_m$, sendo J_m o número total de folhas da iteração.

$$c) \quad \text{Para } j = 1, 2, \dots, J_m \text{ calcule:}$$

$$\gamma_{jm} = \min_{\gamma} \sum_{x^{(i)} \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$$

Ou seja, calcula-se o argumento que minimiza a função perda para cada folha da iteração sobre a predição de f_{m-1} .

d) Atualize $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.

Onde a função $f_m(x)$ é atualizada utilizando-se a função da iteração anterior f_{m-1} e o argumento calculado anteriormente γ_{jm} .

3) O resultado é dado por $\hat{f}(x) = f_M(x)$, calculado na última iteração M .

Por sua vez, o algoritmo do XGBoost apresenta melhorias em relação aos outros modelos de *Gradient Boosting* tais como: penalização inteligente de árvores, redução proporcional dos nós das folhas, método numérico de Newton, além de ser otimizado para computação paralela.

O XGBoost também apresenta regularização, reduzindo, com isso, o sobreajuste e permite executar validação cruzada em cada iteração do processo de *boosting* e, portanto, obter o número ideal de iterações de *boosting* em uma única execução.

4.6. Máquinas de Vetores Suporte

Máquinas de Vetores Suporte - SVM (*Support Vector Machine* – SVM em inglês) é uma técnica que pode ser empregada tanto para classificação quanto para regressão e foi elaborado em um artigo de Boser, Guyon e Vapnik (1992). A SVM é um algoritmo de aprendizado supervisionado, cujo objetivo é classificar um conjunto de dados que são mapeados para um espaço multidimensional usando uma função de núcleo - *kernel*. Nesta função, o limite de decisão no espaço de entrada é representado por um hiperplano em dimensão superior ao espaço original, ou seja, o algoritmo permite a adicionar uma dimensão extra aos dados de modo a encontrar o melhor hiperplano para classificação dos dados (Vapnik *et al.*, 1997 e Saradhi *et al.*, 2005).

O precursor do SVM foi o Classificador de Margem Máxima, aplicado a dados linearmente separáveis. Esse classificador busca encontrar o hiperplano que maximiza a distância entre as classes, chamada de margem, de modo a otimizar o desempenho do modelo em dados fora do treinamento. Ou seja, busca-se um hiperplano que divida os dados em

regiões, de modo que os dados de mesma classe estejam na mesma região do hiperplano. No caso de duas e três dimensões o hiperplano é, respectivamente, uma reta e um plano, por exemplo.

Contudo, os problemas geralmente envolvem dados que não podem ser linearmente separáveis por um hiperplano de distância máxima, apesar de ainda ser possível separá-los por uma superfície linear. Para tratar essa situação, foi desenvolvido o Classificador de Vetor Suporte, uma generalização do Classificador de Margem Máxima. Este algoritmo busca encontrar o hiperplano que classifique corretamente a maioria dos elementos entre as classes e, ao mesmo tempo, maximize a distância entre a maioria dos elementos de diferentes classes, aumentando sua capacidade de generalização e permitindo tratar *outliers*.

Por sua vez, para tratar situações em que não é possível separar os dados utilizando uma superfície linear, foi desenvolvido o modelo de Máquinas de Vetores Suporte. As SVM utilizam uma técnica chamada de truque de *kernel* que mapeia os dados de entrada do espaço original para um novo espaço com dimensões adicionais de modo que eles sejam linearmente separáveis. Segundo Hastie, Tibshirani e Friedman (2017), o processo de otimização para se encontrar o melhor hiperplano que divide os dados envolve somente o cálculo dos produtos escalares entre os dados de treinamento. Ao se utilizar um novo espaço com maiores dimensões, o truque de *kernel* permite o cálculo dos produtos escalares diretamente, sem a necessidade do cálculo das coordenadas dos dados no novo espaço ou mesmo da função que mapeia estes dados no novo espaço.

Já para o caso de regressões, o algoritmo busca o hiperplano cujas margens tenham o maior número possível de dados da amostra de treinamento, minimizando os dados da amostra que estejam do lado de fora das margens, raciocínio oposto ao do processo de classificação.

Em seguida é apresentado como o método de SVM pode ser adaptado para um modelo de regressão linear lidando com generalizações não lineares, apresentado em Hastie, Tibshirani e Friedman (2017). O modelo a ser adaptado é o apresentado a seguir:

$$f(x) = x^T \beta + \beta_0 \quad (17)$$

Para estimar os β , considera-se a minimização de:

$$H(\beta, \beta_0) = \sum_{i=1}^N V(y_i - f(x_i)) + \frac{\lambda}{2} \|\beta\|^2 \quad (18)$$

Onde:

$$V_\epsilon(r) = \begin{cases} 0, & \text{se } |r| < \epsilon \\ |r| - \epsilon, & \text{caso contrário} \end{cases} \quad (19)$$

Ou seja, $V_\epsilon(r)$ é uma função de erro onde são ignorados erros de tamanho menor que ϵ . Analogamente ao classificador de vetor suporte, os dados do lado correto das margens, ou distante delas, são ignorados no processo de otimização. Na regressão, os dados que possuem esses erros são os que apresentam pequenos resíduos. A medida dos erros do vetor de suporte dada pela equação 19 tem caudas lineares além de ϵ e zera as contribuições dos casos com pequenos resíduos.

Se $\hat{\beta}$ e $\widehat{\beta}_0$ são os parâmetros que minimizam H , pode-se demonstrar que a função solução do problema tem a forma:

$$\hat{\beta} = \sum_{i=1}^N (\hat{\alpha}_i^* - \hat{\alpha}_i) x_i \quad (20)$$

$$\hat{f}(x) = \sum_{i=1}^N (\hat{\alpha}_i^* - \hat{\alpha}_i) \langle x, x_i \rangle + \beta_0 \quad (21)$$

Onde $\hat{\alpha}_i^*$ e $\hat{\alpha}_i$ são positivos e resolvem o problema:

$$\min_{\alpha_i, \alpha_i^*} \sum_{i=1}^N (\alpha_i^* + \alpha_i) - \sum_{i=1}^N (\alpha_i^* - \alpha_i) + \frac{1}{2} \sum_{i,i'=1}^N (\alpha_i^* - \alpha_i) (\alpha_{i'}^* - \alpha_{i'}) \langle x_i, x_{i'} \rangle$$

Sujeito a:

$$\begin{aligned} 0 &\leq \alpha_i, \alpha_i^* \leq \frac{1}{\lambda}, \\ \sum_{i=1}^N (\alpha_i^* - \alpha_i) &= 0, \\ \alpha_i \alpha_i^* &= 0. \end{aligned} \quad (22)$$

Dadas essas restrições, somente o subconjunto das soluções cujos valores $(\hat{\alpha}_i^* - \hat{\alpha}_i)$ são diferentes de zero, e seus dados associados, são chamados de vetores de suporte. Esta solução depende dos dados de entrada somente pelo seu produto interno $\langle x_i, x_i \rangle$. Desta forma, pode-se generalizar o método para espaços de maiores dimensões através da definição de diferentes produtos internos em diferentes funções de *kernel*. Percebe-se, que há dois parâmetros na minimização da equação 18, quais sejam: λ e ϵ . ϵ é um parâmetro da função de

perda V_ϵ , que depende da escala de y e, conseqüentemente, de r . Já o λ é um parâmetro tradicional de regularização e pode ser estimado, por exemplo, por validação cruzada.

5. Descrição dos Dados

5.1. Curva de Juros e Variáveis Macroeconômicas

A base de dados utilizada para a projeção da ETTJ foi obtida dos títulos públicos prefixados⁵ (LTN e NTN-F) emitidos pela Secretaria do Tesouro Nacional – STN, além de variáveis macroeconômicas que podem afetar a dinâmica da curva de juros ao longo do tempo.

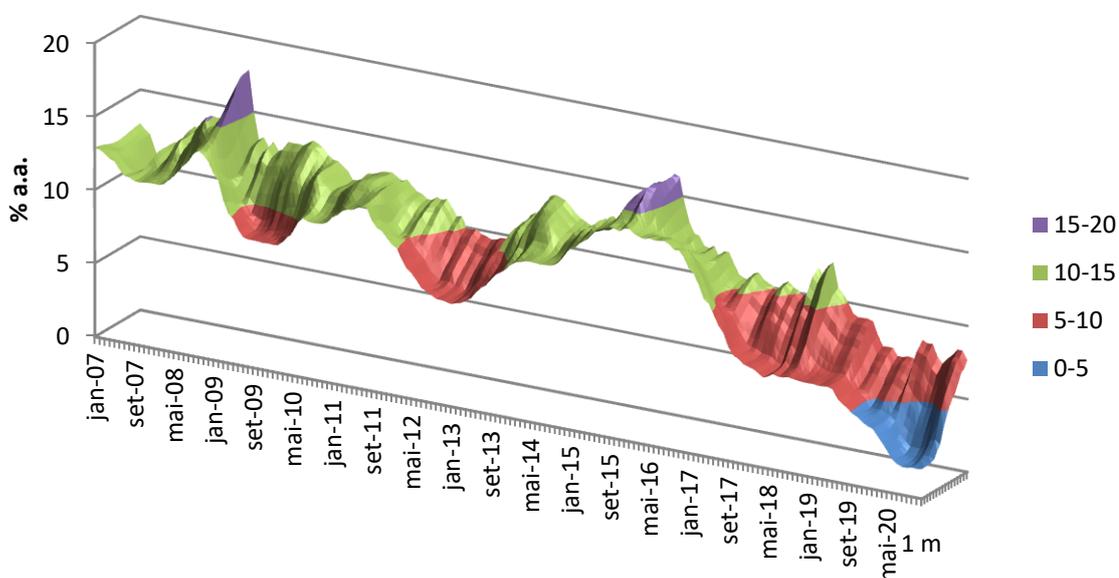
Através das taxas destes títulos no mercado secundário, obtidas na Associação Brasileira das Entidades dos Mercados Financeiros e de Capitais – ANBIMA, foi construído o histórico de curvas zero⁶ para os prazos de 1, 2, 3, 4, 6, 9, 12, 18, 24, 30, 36, 48, 60, 84, 96 e 120 meses, através da técnica de *bootstrapping*⁷. O período analisado vai de janeiro de 2007 a dezembro de 2020. A escolha de janeiro de 2007 para o momento inicial da amostra de dados se deu porque a STN passou a emitir títulos de 10 anos a partir de 2007, permitindo a construção de curvas de juros com prazos mais longos a partir desse ano. As taxas foram geradas diariamente (3516 observações) e, posteriormente, foram construídas taxas mensais por meio da média das taxas diárias (168 observações). A Figura 1 apresenta as curvas de juros observadas.

⁵ Letras do Tesouro Nacional (LTN): são títulos com rentabilidade definida (taxa fixa) no momento da compra. Forma de pagamento: no vencimento. Nota do Tesouro Nacional, série F (NTN-F): são títulos com rentabilidade prefixada, acrescida de juros definidos no momento da compra. Forma de pagamento: semestralmente (juros) e no vencimento (principal).

⁶ É a curva de juros ou estrutura a termo da taxa de juros para fatores de cupom zero, ou seja, curva de juros para estrutura de títulos sem pagamento de cupom entre as datas referenciais.

⁷ A técnica de *bootstrapping* consiste em tratar determinado fluxo de um título como vários fluxos únicos separadamente. Ou seja, cada pagamento de cupom passa a ser tratado como um título sem fluxo.

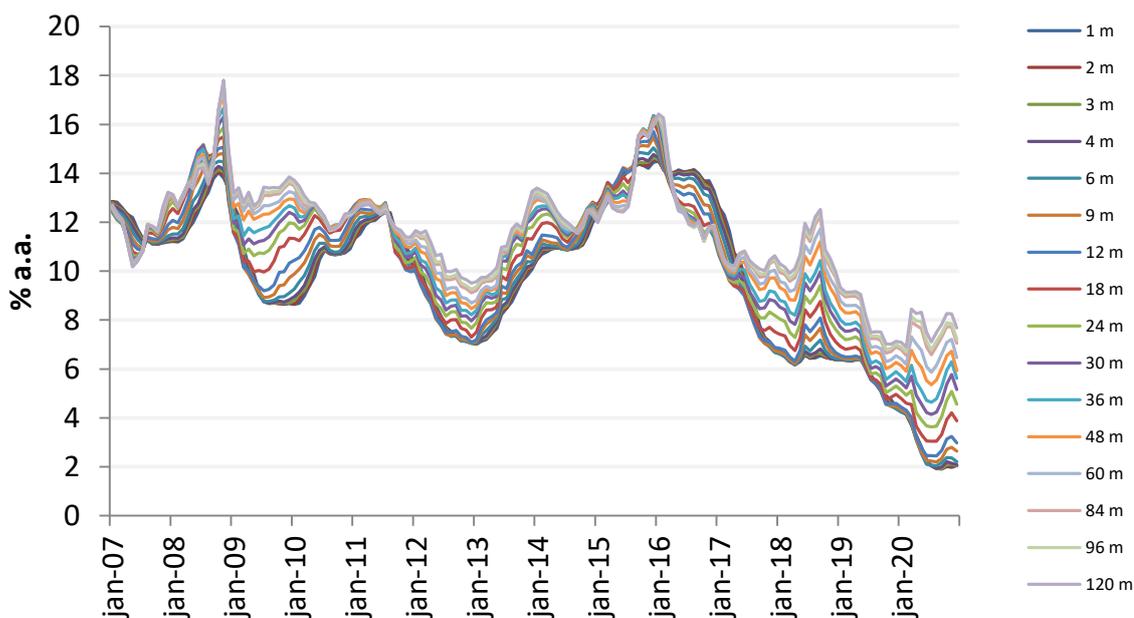
Figura 1 – Evolução da Curva de Juros



Fonte: Elaboração própria, a partir de dados da ANBIMA.

A Figura 2 apresenta a dinâmica das taxas prefixadas para os vértices obtidos na construção das curvas zero, mostrando como a inclinação⁸ da curva varia ao longo do tempo.

Figura 2 – Evolução das Taxas



Fonte: Elaboração própria, a partir de dados da ANBIMA.

⁸ Diferença entre a taxa de maior prazo e a taxa de curto prazo.

Ainda com relação aos dados das curvas de juros, a Tabela 1 apresenta as estatísticas descritivas da ETTJ para a amostra considerada. Nota-se que a curva de juros média apresenta inclinação positiva, em linha com os resultados apresentados na literatura. As volatilidades (desvio padrão) das taxas se reduzem à medida que os prazos ficam mais longos. Pela amplitude das taxas, apresentada pelas taxas máximas e mínimas, percebe-se a significativa redução no nível da curva durante o período analisado. Por fim, as autocorrelações (três últimas colunas) refletem a persistência da dinâmica das taxas.

Tabela 1 – Estatísticas Descritivas

Prazo	Média % a.a.	Desvio Padrão % a.a.	Mínimo % a.a.	Máximo % a.a.	$\rho(1)$	$\rho(12)$	$\rho(24)$
1 m	9,69	3,22	1,92	14,50	0,97	0,44	0,04
2 m	9,69	3,23	1,91	14,59	0,98	0,45	0,05
3 m	9,69	3,23	1,91	14,64	0,98	0,45	0,05
4 m	9,70	3,24	1,94	14,77	0,98	0,45	0,05
6 m	9,74	3,24	2,04	15,04	0,98	0,45	0,05
9 m	9,83	3,22	2,20	15,47	0,98	0,46	0,06
12 m	9,94	3,19	2,44	15,72	0,97	0,46	0,07
18 m	10,22	3,06	3,04	16,15	0,97	0,47	0,09
24 m	10,48	2,91	3,63	16,27	0,97	0,47	0,10
30 m	10,69	2,77	4,15	16,35	0,97	0,47	0,11
36 m	10,86	2,65	4,64	16,66	0,97	0,47	0,11
48 m	11,09	2,48	5,35	17,29	0,96	0,45	0,11
60 m	11,22	2,36	5,86	17,61	0,96	0,43	0,10
84 m	11,41	2,20	6,56	17,73	0,95	0,41	0,10
96 m	11,47	2,16	6,70	17,76	0,95	0,40	0,10
120 m	11,57	2,10	6,90	17,80	0,95	0,38	0,10

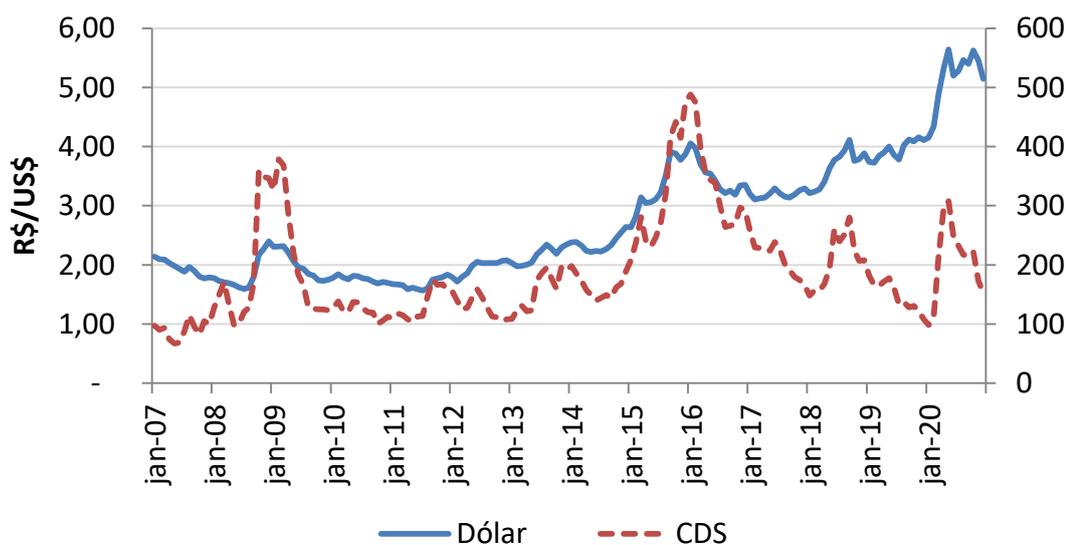
Fonte: Elaboração própria.

Além da série histórica das curvas de juros, este trabalho também utilizou variáveis macroeconômicas cujas dinâmicas estão correlacionadas com a dinâmica dos vértices das curvas de juros. As variáveis utilizadas são a taxa SELIC, o Índice Geral de Preços ao Consumidor – IPCA, a expectativa de inflação (IPCA) doze meses à frente, o Dólar, e o *Credit Default Swap*⁹ – CDS de 5 anos do Brasil.

O nível da curva de juros é dado pela taxa SELIC. Adicionalmente, os movimentos das taxas de curto prazo da curva de juros estão correlacionados com o movimento da taxa básica de juros da economia. Por sua vez, as taxas para prazos longos normalmente estão correlacionadas com a variação do Dólar, em função dos investidores estrangeiros que atuam nesse nicho de mercado, e com a variação do CDS, que é uma forma de mensurar o risco país. Por sua vez, tanto o IPCA, quanto a expectativa de inflação, afetam o prêmio de risco presente na curva prefixada, afetando sua inclinação. Essas variáveis macroeconômicas encontram-se nas figuras 3 a 5 a seguir, para o período de análise deste trabalho.

Percebe-se, comparando as figuras dessas variáveis com a Figura 2 que, nos momentos em que há maior percepção de risco, como quando ocorre o aumento da taxa de câmbio, do CDS ou da expectativa de inflação, também ocorre um movimento de aumento nas taxas de juros de longo prazo. Enquanto os vértices de curto prazo tendem a acompanhar o movimento da SELIC, que, por sua vez, reage ao aumento da inflação.

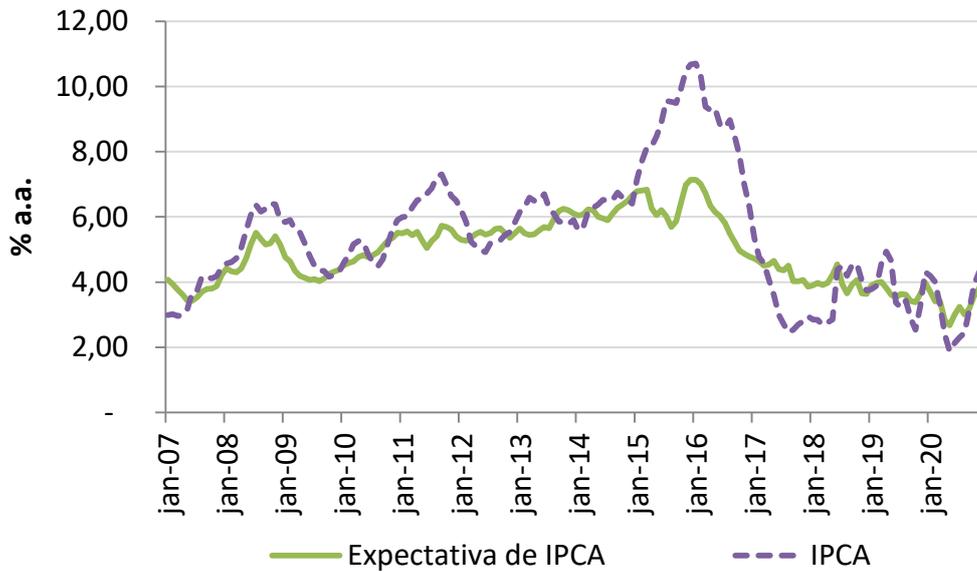
Figura 3 – Evolução do Dólar e do CDS Brasil de 5 anos



Fonte: Dólar - Banco Central do Brasil; CDS - <https://www.investing.com>. Elaboração própria.

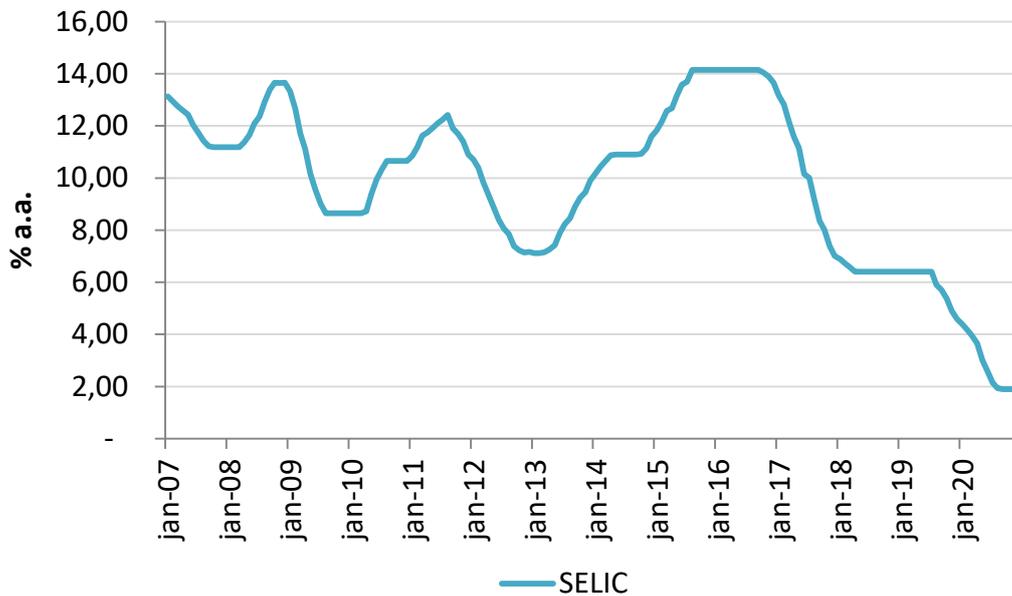
⁹ *Credit default swap* é um contrato de swap que remunera o portador quando ocorre o default da instituição especificada no contrato.

Figura 4 – Evolução do IPCA e da Expectativa de IPCA 12 meses a frente



Fonte: Elaboração própria, a partir de dados do Banco Central do Brasil.

Figura 5 – Evolução da Taxa SELIC



Fonte: Elaboração própria, a partir de dados do Banco Central do Brasil.

5.2. Evolução da Economia e da Curva de Juros de 2007 a 2020

A dinâmica das curvas de juros ao longo dos anos e sua relação com a economia brasileira é descrita nos próximos parágrafos, contemplando o período de 2007 a 2020.

O ano de 2007 iniciou-se de forma otimista com relação a 2006, em função de um cenário de expansão da atividade econômica em um contexto de estabilidade dos preços e de aumento de renda. Ao longo do ano ocorreu flexibilização da política monetária, com a redução da taxa básica de juros com a SELIC se estabilizando em 11,25% a.a. no último trimestre. Com esse cenário benigno, a curva de juros apresentava-se negativamente inclinada, com a ponta longa apresentando valores inferiores à taxa básica de juros. Esse movimento se inverteu ao longo do segundo semestre em função do aumento da volatilidade no mercado externo decorrente da crise de crédito na economia norte-americana iniciada no segmento de *subprime* do mercado hipotecário. Desta forma, ao final do ano, a curva de juros passou a ter inclinação positiva em função do aumento da percepção de risco no cenário externo.

Já o ano de 2008 foi marcado por grande volatilidade decorrente da crise financeira internacional, decorrente da deterioração do cenário econômico mundial, e devido aos riscos inflacionários no cenário doméstico. Desta forma, houve ao longo do ano um endurecimento da política monetária com o aumento na taxa básica de juros e uma deterioração da curva de juros. Em outubro o ambiente era de alta aversão do risco e, com isso, ocorreu um forte aumento nas taxas da curva de juros brasileira, especialmente nas taxas de prazo mais longo, forçando a STN a efetuar leilões de compra e venda de títulos, de forma a criar referência de preços para o mercado e prover liquidez aos investidores que buscavam se desfazer de suas posições no Brasil com o objetivo de reduzir perdas em seus países de origem, além de ajustar suas exposições aos países em suas carteiras de investimentos. Nos últimos dois meses do ano, a inflação apresentava-se inferior à estimativa de mercado e os dados mostravam recuo da atividade econômica. Externamente houve uma melhora na percepção de risco com relação à crise financeira internacional. Com isso a SELIC manteve-se estável no final do ano e houve uma queda significativa nos prêmios de risco da curva de juros doméstica.

O ano de 2009, por sua vez, foi marcado pela solidez dos fundamentos econômicos no Brasil, inclusive com a concessão do grau de investimento por uma agência de classificação de risco. O Brasil foi um dos últimos países a sentir os efeitos da crise internacional e foi um dos primeiros a mostrar recuperação consistente da economia. A condução da política econômica de forma responsável em anos anteriores permitiu a tomada de diversas medidas de caráter anticíclico, criando condições para a retomada do crescimento do país. Em função do cenário benigno de inflação e, em conjunto a medidas de estímulo à economia, houve queda na SELIC ao longo de todo o ano de 2009. Em julho, o Banco Central encerrou o ciclo

de flexibilização monetária, com o registro do menor patamar da SELIC da história (8,75% a.a.). Por outro lado, no mercado doméstico, apesar da inflação encontrar-se compatível com a meta da autoridade monetária, as taxas de longo prazo aumentaram devido ao otimismo com o cenário internacional, que poderia fazer com que o Brasil sofresse pressões inflacionárias do resto do mundo.

Já o início de 2010 foi marcado por uma robusta expansão econômica, sustentada pelo consumo interno e refletindo, em parte, os efeitos das medidas de estímulos implantadas para combater as consequências da crise financeira em 2009. Esse cenário de expansão econômica começou a gerar uma deterioração das expectativas de inflação que levou ao aumento da taxa básica de juros ao longo do ano, encerrando o ciclo de aperto monetário em julho, com a SELIC em 10,75%. Outros fatores importantes ao longo do ano foram o agravamento das condições fiscais na Europa e o consequente aumento da volatilidade nos mercados financeiros internacionais. Esses fatores acabaram gerando volatilidade nas taxas mais longas da curva de juros doméstica.

Já o ano de 2011 pode ser dividido em dois momentos. O primeiro momento pode ser resumido principalmente pelos efeitos decorrentes da pressão que havia sobre os níveis de preços. A incerteza sobre o processo inflacionário criou expectativas sobre a elevação da taxa básica de juros, fazendo com que a curva doméstica de juros apresentasse um aumento nas taxas mais curtas. Além disso, o Banco Central deu início a um novo ciclo de aperto monetário que durou até agosto deste ano, com a SELIC atingindo o valor de 12,5% a.a.. Já o segundo momento teve início com um novo ciclo de flexibilização monetária de forma a combater os efeitos da deterioração do cenário externo, decorrente dos desdobramentos da crise internacional e causada por dúvidas quanto à solvência da dívida dos países da Zona do Euro. Essa crise teve efeitos na economia doméstica, iniciando um processo de desaceleração da economia com uma consequente convergência da inflação para a sua meta no médio prazo. Com a desaceleração da economia, além da redução na taxa básica de juros, ocorreu um aumento do deslocamento da curva de juros para baixo ao longo do segundo semestre. Apesar do deslocamento para baixo nas taxas de curto e de médio prazo, no final do ano, ainda havia pressão nas taxas de longo prazo, em função da expectativa de que a economia iria se recuperar e apresentar um aumento no ritmo de crescimento, novamente.

Em 2012, ocorre nova redução do nível das curvas de juros, decorrentes do esforço do governo brasileiro em combater os efeitos da crise na Zona do Euro e da desaceleração do crescimento da economia doméstica, através de, entre outras ações, de uma ampla

flexibilização monetária que fez com que a SELIC recuasse em 3,57 pontos percentuais ao longo do ano, atingindo o mínimo histórico de 7,25% a.a.. Apesar dessa forte redução na taxa básica da economia, a inflação encerrou o ano em nível inferior ao de 2011. Essa flexibilização monetária fez com que a curva de juros apresentasse forte inclinação negativa em seus prazos mais curtos e uma forte inclinação positiva nos prazos mais longos com a perspectiva de recuperação da economia, decorrentes das medidas de estímulo que foram tomadas ao longo do ano.

O ano de 2013 foi pautado, no plano externo, pela expectativa de retirada dos estímulos monetários postos em prática para estimular a economia americana. Já no ambiente doméstico, o ano começou com uma expectativa otimista quanto ao crescimento econômico, sobretudo devido aos estímulos feitos no ano anterior para reverter a desaceleração econômica. Contudo, a retomada da atividade econômica foi menos intensa que a esperada e as pressões inflacionárias começaram a se fazer presentes, com a elevação no nível de preços. Desta forma, a expectativa de crescimento do PIB, tanto para 2013 quanto para 2014, começaram a se deteriorar fortemente a partir do segundo semestre. Com a inflação próxima ao limite oficial, o Banco Central iniciou o ciclo de aperto monetário em abril. Já no âmbito fiscal, o superávit primário indicava perda de força em decorrência da ausência da recuperação econômica e da política de desoneração tributária sobre o setor produtivo. Nesse contexto, a autoridade monetária intensificou o ritmo de aumento da SELIC que terminou o ano em 10% a.a.. A curva de juros apresentou praticamente a mesma inclinação ao longo do ano, com uma abertura em torno de 3 pontos percentuais entre as taxas de longo prazo e as de curto prazo, se movendo de forma paralela, para cima, à medida que a taxa básica de juros aumentava.

Em 2014, a inflação continuou pressionada ao longo do ano, apesar da política monetária restritiva, que elevou a taxa básica de juros a 11,75% a.a. ao final do ano. Adicionalmente, a percepção de risco decorrente da crise fiscal continuou aumentando, devido à queda na arrecadação tributária, oriunda da desaceleração econômica, das desonerações tributárias e da maior pressão sobre os gastos. Nesse contexto, ainda no primeiro semestre, já havia uma piora na expectativa de desempenho econômico com relação às previsões feitas no início do ano, tanto para 2014 quanto para 2015. Outro fator que gerou volatilidade foi a incerteza decorrente do processo eleitoral. Com relação à curva de juros, ao longo do ano, as taxas mais curtas acompanharam o aperto monetário e aumentaram, principalmente no primeiro trimestre e após as eleições, em outubro. Entre abril e outubro as

taxa de longo prazo começaram a ceder, mas após as eleições, e em virtude da incerteza com relação à nova equipe econômica e às possíveis medidas de ajuste para os próximos anos houve um aumento da percepção de risco, com aumento da volatilidade e do nível das taxas de longo prazo. Apesar de todo esse movimento a curva de juros terminou o ano com uma pequena inclinação positiva, decorrente do elevado nível de juros no curto prazo.

No ano de 2015, apesar do aperto monetário, que levou o Banco Central a elevar a taxa SELIC a 14,25% a.a. em julho, a inflação seguiu pressionando a economia, sobretudo devido à correção tarifária que ocorreu no início do ano e ao ajuste na taxa de câmbio. Com isso, a expectativa de inflação para 2015 foi aumentando sistematicamente ao longo do ano. Adicionalmente, o cenário fiscal se deteriorou rapidamente com a retração significativa da atividade econômica. Já em setembro, o PIB acumulava queda de aproximadamente 4,5%. Nesse contexto fiscal, o governo encaminhou ao Congresso Nacional em agosto, o projeto de Lei Orçamentária Anual para 2016 prevendo um déficit fiscal de R\$ 30 bilhões e, na sequência, o Brasil teve sua nota de crédito rebaixada, perdendo um dos graus de investimento obtido em 2008. Desta forma, a partir de setembro, com o aumento da percepção de risco, as taxas longas voltaram a ficar em um nível mais elevado e houve um aumento da volatilidade das taxas, o que levou a STN a intervir no mercado de títulos públicos novamente através de leilões extraordinários de compra e venda, de forma a fornecer referência de preços aos investidores. No final do ano, ainda houve nova volatilidade nas taxas de longo prazo decorrente da perda do grau de investimento por outra agência de classificação de risco.

O ano de 2016 começou com um agravamento do cenário econômico e fiscal, mas, a partir do processo de impeachment, identificou-se uma recuperação dos indicadores de confiança, bem como a redução do risco Brasil e apreciação do Real. Apesar da expectativa inicial de um cenário mais otimista de recuperação da atividade econômica, este cenário acabou não ocorrendo e a recuperação se deu de forma mais lenta que a esperada, com o país apresentando, assim como em 2015, uma significativa redução do PIB real. A inflação, por sua vez, voltou para o intervalo de cumprimento das metas de inflação, fechando o ano em 6,3%, ante uma inflação de 10,7% no ano anterior. O controle da inflação levou o Banco Central a começar um novo ciclo de flexibilização monetária em outubro de 2016, com a SELIC atingindo 13,75% a.a. em dezembro.

Com relação ao comportamento da curva de juros em 2016, até março, ela apresentava inclinação positiva. A partir de abril, a curva de juros passou a ter inclinação negativa, com as taxas de prazos mais longos inferiores às taxas de médio e de curto prazo. Isso ocorreu devido

ao aumento da credibilidade na condução da política monetária e com os agentes acreditando na capacidade do Banco Central de levar a inflação a convergir para a sua meta. Adicionalmente, destaca-se a aprovação da Emenda Constitucional nº 95, de 15 de dezembro de 2016, que instituiu o Novo Regime Fiscal. Esta emenda, conhecida como Teto dos Gastos, limitou a taxa de crescimento das despesas da União para os próximos 20 anos, criando uma âncora para a política fiscal e melhorando as expectativas do mercado, que se traduziram em uma redução da percepção de risco país.

O ano de 2017 se iniciou com um cenário de recuperação lenta e gradual da atividade econômica. Por sua vez, o enfraquecimento da atividade nos últimos dois anos impactou expressivamente os preços, com a perspectiva de que a inflação, ao longo de 2017, iria convergir para um valor próximo ao centro da meta, além de favorecer as projeções de inflação para os próximos anos. Esse contexto permitiu que o Banco Central continuasse o processo de redução da taxa básica de juros iniciado em 2016, saindo de um patamar de 13,5% a.a. para 7% a.a. ao final de 2017. Desta forma, observou-se uma significativa redução das taxas de juros de curto prazo ao longo do ano, refletindo, inclusive, a expectativa de cortes na taxa SELIC no curto prazo. Por outro lado, as taxa de juros de longo prazo não acompanharam toda essa redução que ocorreu nos prazos mais curtos, uma vez que as taxas mais longas refletem preocupações com a economia no longo prazo, tais como a consolidação fiscal decorrente de medidas como a reforma da Previdência e as eleições de 2018.

Já o ano de 2018, em um contexto de inflação sob controle (o IPCA fechou 2018 em 3,75%, abaixo da meta de 4,5%), apresentou um período de política monetária expansionista durante quase todo o ano, com a taxa SELIC nas mínimas históricas, no valor de 6,5% a.a.. A taxa básica de juros da economia nesse nível mínimo ajudou a mitigar os efeitos adversos relacionados ao aumento da dívida pública decorrente do ambiente de déficit fiscal que ocorria desde 2014. Nesse cenário, a economia brasileira seguiu o processo de recuperação, ainda que de forma mais lenta do que o esperado inicialmente. Com relação à curva de juros, as taxas de curto prazo pouco se alteraram uma vez que a SELIC se manteve estável durante quase todo o ano. Por sua vez, tanto as taxas de médio prazo, quanto as de longo prazo, refletiram ao longo do ano o aumento na percepção de risco decorrente de eventos tais como a greve dos caminhoneiros e de incertezas quanto às reformas fiscais, atingindo valores máximos em junho e em setembro. Apesar dessa volatilidade, a inclinação da curva de juros cedeu no final do ano, com a redução do diferencial entre as taxas de longo prazo e as de curto prazo.

No ano de 2019, com as expectativas inflacionárias ancoradas e com a percepção de que a recuperação econômica ainda seria lenta, o Banco Central continuou a política inflacionária expansionista, com a taxa básica de juros, SELIC, atingindo seu menor valor histórico ao final do ano, no valor de 4,5% a.a.. Além do cenário benigno de inflação e dos níveis historicamente baixos da SELIC, o Brasil promoveu importantes passos rumo à consolidação fiscal. A reforma da Previdência foi aprovada no Congresso Nacional, ocorreram os leilões de excedentes da cessão onerosa de petróleo e houve redução de ativos da União com bancos públicos, com importantes efeitos sobre o endividamento público. Por sua vez, a atividade econômica ainda permaneceu em um ritmo inferior ao projetado inicialmente, devido principalmente a incertezas relacionadas a reformas estruturais e ao cenário externo conturbado. Nesse contexto, as taxas de juros de curto prazo seguiram a redução da taxa SELIC ao longo do ano. Adicionalmente, além das taxa de longo prazo também estarem em seus mínimos históricos, com a melhora da economia e com a aprovação da reforma da Previdência, o prêmio de risco, definido como a diferença entre as taxas de longo e de curto prazo, fechou ainda mais se comparado com o de 2018.

Por fim, 2020 foi um ano de muitos desafios em virtude da pandemia de COVID-19 e com impactos sem precedentes na economia mundial. No Brasil, o combate à pandemia ampliou significativamente os gastos públicos e a adoção de medidas de isolamento social provocou choques sobre o mercado de trabalho e na atividade econômica, provocando uma queda no PIB de 4,5%, mesmo com a economia apresentando recuperação a partir de junho. Por sua vez, a inflação manteve-se controlada durante todo o ano devido ao enfraquecimento da atividade econômica. Medidas de estímulo monetário, fiscal e regulatório foram tomadas para proteger as empresas e a vida e o emprego das pessoas. Desta forma, a taxa básica de juros voltou a atingir mínimos históricos, de acordo as com medidas de estímulo monetário, atingindo um mínimo de 2% a.a. a partir de agosto. Em março, em decorrência das incertezas provocadas pela pandemia, as taxas de juros abriram de forma repentina e a STN teve que atuar novamente com um programa de compra e venda simultânea de títulos, com o objetivo de fornecer referência de preços ao mercado. As taxas longas, por sua vez, ao longo do ano, não acompanharam a redução da taxa básica de juros devido ao risco fiscal e das incertezas decorrentes da pandemia, deixando a curva de juros brasileira uma das mais inclinadas do mundo.

6. Resultados

6.1. Procedimentos de Previsão

Para se efetuar as projeções fora da amostra, inicialmente, os modelos são estimados utilizando a primeira metade da amostra, ou seja, uma janela de estimação de 74 meses. A primeira janela de estimação corresponde ao período de 08/2007 a 09/2013. Na sequência, os modelos são estimados novamente, incluindo o próximo mês na amostra¹⁰. Os parâmetros são atualizados até a última amostra de 12/2019. Para cada intervalo de estimação são feitas as projeções da curva de juros¹¹.

As projeções são feitas para os seguintes prazos: 1, 3, 6 e 12 meses. Os modelos de aprendizado de máquina são comparados ao modelo de Passeio Aleatório, que é usado como referência na literatura de projeção de curva de juros.

As covariadas utilizadas para o treinamento foram, em um primeiro momento, os vértices das curvas de juros e suas defasagens até a sexta ordem. Posteriormente, a análise foi feita adicionando as variáveis macroeconômicas e suas defasagens, também até a sexta ordem. Adicionalmente, o treinamento e as projeções foram feitas com os dados transformados para a diferença dos logaritmos das covariadas. Por sua vez, a busca pelos hiperparâmetros ótimos foi feita com a técnica de otimização Bayesiana para todos os modelos, exceto o modelo LASSO, onde foi utilizada a técnica de *random search*.

6.2. Procedimentos de Avaliação dos Resultados

Para a avaliação do período fora da amostra, é calculada, para cada horizonte de previsão e prazo de maturidade dos vértices da curva de juros, a Raiz do Erro Quadrático Médio - REQM.

$$REQM_c^{\tau_i} = \sqrt{\frac{1}{P_i} \sum_{t=1}^{P_i} (\hat{y}_{t+h|t,c}^{\tau_i} - y_{t+h}^{\tau_i})^2} \quad (23)$$

¹⁰ Neste trabalho, utilizar uma janela crescente de dados para treino apresentou resultados superiores a uma janela móvel de tamanho fixo.

¹¹ As projeções foram feitas em um programa desenvolvido em R. No programa foi utilizada a interface Machine Learning in R – MLR, que apresenta os modelos utilizados neste trabalho.

Por sua vez, a precisão das projeções em comparação com o Passeio Aleatório é avaliada pela Raiz do Erro Quadrático Médio Relativo – ReREQM:

$$ReREQM_c^{\tau_i} = \frac{REQM_c^{\tau_i}}{REQM_{PA}^{\tau_i}} \quad (24)$$

De forma similar a Araujo (2011) e Hordal *et al.* (2006), os desempenhos dos modelos são resumidos pelo Traço da Raiz do Erro Quadrático Médio – TREQM, proposto por Christoffersen e Diebold (1998):

$$TREQM_c = \sqrt{\frac{1}{N} \frac{1}{P_i} \sum_{i=1}^N \sum_{t=1}^{P_i} (\hat{y}_{t+h|t,c}^{\tau_i} - y_{t+h}^{\tau_i})^2} \quad (23)$$

Por fim, para testar se as previsões fora da amostra entre os modelos de Aprendizado de Máquina e o Passeio Aleatório são estatisticamente semelhantes ou não, é utilizado o teste de Diebold-Mariano¹² - DM, que compara as previsões entre dois modelos diferentes.

Assume-se, portanto, para essa avaliação, que existem duas séries de previsão e, conseqüentemente, duas funções de perdas $g(y_t, \hat{y}_{it})$. A função de perda, em geral, está associada aos erros de previsão, conforme as equações (24) e (25):

$$g(y_t, \hat{y}_{it}) = g(e_{it}) \quad (24)$$

$$e_{it} = \hat{y}_{it} - y_t, \quad i = 1, 2 \quad (24)$$

Com as funções de perda, é definida a diferença estimada entre os dois modelos de previsão:

$$d_t = g(e_{1t}) - g(e_{2t}) \quad (25)$$

Dessa forma, a hipótese nula do teste, a qual considera que a função de perda dos dois modelos gera previsões estatisticamente iguais, é dada por:

$$H_0: E[d_t] = 0 \quad \forall t \quad (26)$$

¹² Diebold e Mariano (1995) e Diebold (2013).

Por sua vez, as hipóteses alternativas podem ser dadas por:

$$H_1: E[d_t] > 0 \quad (27)$$

em que o segundo modelo possui melhor modelo preditivo que o primeiro modelo.

Ou:

$$H_1: E[d_t] < 0 \quad (28)$$

na qual o primeiro modelo possui melhor modelo preditivo que o segundo modelo.

Ou:

$$H_1: E[d_t] \neq 0 \quad (28)$$

para a alternativa bivariada, onde as previsões são estatisticamente diferentes.

A estatística do teste t de Diebold-Mariano é dada por:

$$t_{DM} = \frac{\bar{d}}{\hat{\sigma}_d} \sim N(0,1) \quad (29)$$

onde $\hat{\sigma}_d$ é o erro padrão de \bar{d} . Os autores do teste definem $\hat{\sigma}_d = \sqrt{T^{-1}\hat{g}(0)}$, onde $\hat{g}(0)$ é um estimador consistente da variância do diferencial de perda no espectro de frequência zero e T é o prazo de projeção.

Por fim, Diebold (2013) estabelece as condições para a condução do teste, quais sejam:

$$E[d_{t+1}] = \mu_d \quad \forall t \quad (30)$$

$$Cov(d_t, d_{t-j}) = \gamma_j \quad (31)$$

$$0 < var(d_t) = \sigma^2 \quad (32)$$

6.3. Resultados dos Modelos de Aprendizado de Máquina

Os desempenhos dos modelos são analisados considerando os resultados do Traço da Raiz do Erro Quadrático Médio Relativo, TReREQM, para cada prazo da estrutura a termo da taxa de juros e para cada horizonte de projeções. A segunda coluna das tabelas apresenta o Traço da Raiz do Erro Quadrático Médio, TREQM, do Passeio Aleatório. As colunas

restantes da tabela apresentam o TReREQM dos outros modelos em relação ao Passeio Aleatório conforme a equação 24.

O TReREQM menor do que 1 indica que o modelo supera o Passeio Aleatório, enquanto um número maior do que 1 indica que o Passeio Aleatório é superior.

Os resultados são apresentados nas Tabelas 2 e 9. As Tabelas 2 a 5 apresentam os resultados dos modelos treinados somente com os dados históricos das curvas de juros, enquanto as Tabelas 6 a 9 apresentam os resultados dos modelos treinados com as variáveis macroeconômicas em conjunto com os dados das curvas de juros.

Com relação aos modelos treinados somente com os dados das curvas de juros, as projeções 1 mês à frente dos modelos testados superaram o Passeio Aleatório na maioria das maturidades, conforme mostra a Tabela 2.

Neste caso, nos modelos K Vizinhos mais Próximos e nas Floresta Aleatórias, todos os vértices apresentaram resultado superior ao Passeio Aleatório. Já o modelo Lasso foi superior ao PA até o vértice de 24 meses, enquanto o SVM foi superior até o vértice de 36 meses, com a exceção do vértice de 6 meses na parte curta da curva de juros. Por sua vez, o modelo XGB foi superior ao PA na parte curta da curva de juros, até o vértice de 12 meses, e em alguns vértices na parte intermediária da curva, nos vértices de 24, 48 e 60 meses, conforme Tabela 2 a seguir.

Tabela 2 – Resultados Curvas de Juros, horizonte de 1 mês

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	0,310	0,690	0,983	0,587	0,563	0,578
2 m	0,304	0,709	0,989	0,614	0,609	0,671
3 m	0,303	0,733	0,989	0,588	0,667	0,739
4 m	0,311	0,782	0,990	0,628	0,732	0,800
6 m	0,326	0,854	0,991	0,749	0,871	1,051
9 m	0,362	0,912	0,994	0,829	0,947	0,997
12 m	0,390	0,948	0,995	0,909	0,925	0,989
18 m	0,435	0,971	0,998	0,937	1,052	0,998
24 m	0,460	0,975	0,999	0,954	0,971	0,999
30 m	0,484	0,977	1,001	0,953	1,077	0,997
36 m	0,498	0,975	1,001	0,942	1,039	0,997
48 m	0,521	0,971	1,002	0,990	0,999	1,005
60 m	0,537	0,972	1,003	0,993	0,998	1,008
84 m	0,553	0,970	1,004	0,993	1,017	1,003
96 m	0,557	0,969	1,004	0,991	1,059	1,010
120 m	0,560	0,967	1,004	0,983	1,068	1,005

Fonte: Elaboração própria.

Contudo, nos demais prazos de projeção, de 3, 6 e 12 meses, nenhum modelo apresentou resultado superior ao Passeio Aleatório de forma sistemática.

Para o prazo de 3 meses, de acordo com a Tabela 3, os modelos foram superiores ao PA somente nos vértices de curto prazo, principalmente até o vértice de 12 meses, sendo que os modelos LASSO e Floresta Aleatória apresentaram o melhor desempenho até os vértices de 18 e 24 meses, respectivamente.

Já para o prazo de projeção de 6 meses, conforme apresentado na Tabela 4, os modelos só foram superiores ao PA na parte curta da curva, até o vértice de 12 meses, nos modelos LASSO, Floresta Aleatória e SVM, exceto no vértice de 6 meses para este último, enquanto os modelos kNN e XGB só foram superiores até o vértice de 4 meses.

Com relação ao prazo de 12 meses, de acordo com a Tabela 5, o PA foi superior aos modelos na maioria dos vértices, com algumas exceções nos modelos LASSO, Floresta Aleatória e SVM.

Tabela 3 – Resultados Curvas de Juros, horizonte de 3 meses

Prazo	REQM	TREQM				
	PA	Knn	LASSO	RF	XGB	SVM
1M	0,614	0,741	0,987	0,595	0,577	0,643
2M	0,612	0,778	0,992	0,627	0,652	0,712
3M	0,613	0,814	0,992	0,635	0,714	0,785
4M	0,618	0,855	0,992	0,684	0,769	0,776
6M	0,629	0,917	0,992	0,770	0,912	0,996
9M	0,661	0,977	0,994	0,853	1,040	0,991
12M	0,688	1,019	0,996	0,917	0,998	0,989
18M	0,745	1,053	0,999	0,978	1,075	1,000
24M	0,779	1,057	1,001	0,999	1,108	1,000
30M	0,816	1,056	1,003	1,016	1,136	1,002
36M	0,830	1,056	1,010	1,023	1,161	1,009
48M	0,871	1,041	1,006	1,032	1,071	1,005
60M	0,892	1,039	1,007	1,053	1,102	1,006
84M	0,916	1,031	1,008	1,027	1,066	1,010
96M	0,923	1,029	1,009	1,030	1,078	1,011
120M	0,932	1,023	1,010	1,025	1,047	1,012

Fonte: Elaboração própria.

Tabela 4 – Resultados Curvas de Juros, horizonte de 6 meses

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	1,074	0,874	0,983	0,706	0,731	0,744
2 m	1,061	0,917	0,996	0,732	0,811	0,794
3 m	1,061	0,943	0,996	0,754	0,868	0,844
4 m	1,064	0,977	0,996	0,789	0,900	0,844
6 m	1,069	1,016	0,996	0,874	1,064	1,009
9 m	1,086	1,059	0,996	0,939	1,118	0,995
12 m	1,099	1,096	0,997	0,998	1,092	0,992
18 m	1,122	1,132	1,009	1,034	1,130	1,011
24 m	1,153	1,118	1,003	1,055	1,145	1,001
30 m	1,189	1,110	1,007	1,041	1,128	1,006
36 m	1,205	1,103	1,009	1,063	1,124	1,005
48 m	1,241	1,094	1,012	1,055	1,093	1,012
60 m	1,265	1,088	1,014	1,069	1,099	1,013
84 m	1,293	1,080	1,016	1,045	1,108	1,017
96 m	1,306	1,076	1,017	1,048	1,072	1,021
120 m	1,319	1,070	1,018	1,053	1,093	1,023

Fonte: Elaboração própria.

Tabela 5 – Resultados Curvas de Juros, horizonte de 12 meses

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	1,895	1,108	1,004	0,916	1,021	0,945
2 m	1,891	1,125	1,005	0,931	1,074	0,965
3 m	1,886	1,127	1,004	0,953	1,120	1,019
4 m	1,883	1,142	1,004	0,961	1,155	1,017
6 m	1,871	1,134	1,003	1,032	1,254	1,095
9 m	1,847	1,147	1,001	1,081	1,293	1,012
12 m	1,814	1,165	0,999	1,110	1,199	0,999
18 m	1,753	1,170	0,998	1,107	1,252	1,003
24 m	1,720	1,155	1,000	1,121	1,185	0,996
30 m	1,715	1,141	1,004	1,091	1,154	1,004
36 m	1,711	1,131	1,008	1,106	1,148	1,004
48 m	1,716	1,120	1,014	1,073	1,127	1,015
60 m	1,723	1,114	1,017	1,074	1,060	1,017
84 m	1,750	1,105	1,022	1,068	1,140	1,024
96 m	1,766	1,102	1,024	1,059	1,075	1,030
120 m	1,780	1,098	1,027	1,069	1,116	1,035

Fonte: Elaboração própria.

Já com relação aos modelos treinados com os dados das curvas de juros e as variáveis macroeconômicas, os resultados foram qualitativamente semelhantes aos exercícios

anteriores, com as projeções 1 mês à frente dos modelos testados, superando o Passeio Aleatório também na maioria das maturidades.

Neste exercício, somente o modelo kNN apresentou resultado superior ao PA em todos os vértices. O modelo Floresta Aleatória só não foi superior ao modelo PA nos vértices intermediários, de 36, 48 e 60 meses. O modelo XGB apresentou resultados ligeiramente piores que a Floresta Aleatória, com resultados inferiores ao PA entre os vértices de 24 a 60 meses. Já o modelo LASSO foi superior até o vértice de 24 meses e somente o modelo SVM apresentou mais vértices com resultado inferior ao PA para projeções com o horizonte de 1 mês, conforme pode ser observado na Tabela 6 a seguir:

Tabela 6 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 1 mês

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	0,310	0,685	0,990	0,585	0,523	0,863
2 m	0,304	0,680	0,989	0,610	0,611	1,165
3 m	0,303	0,703	0,989	0,604	0,655	1,336
4 m	0,311	0,754	0,990	0,660	0,694	1,403
6 m	0,326	0,829	0,991	0,740	0,813	1,492
9 m	0,362	0,898	0,994	0,805	0,918	0,990
12 m	0,390	0,939	0,995	0,902	0,959	0,992
18 m	0,435	0,968	0,998	0,949	0,979	0,997
24 m	0,460	0,978	0,999	0,961	1,037	0,999
30 m	0,484	0,982	1,001	0,988	1,089	0,999
36 m	0,498	0,981	1,001	1,014	1,031	1,000
48 m	0,521	0,978	1,002	1,014	1,001	1,002
60 m	0,537	0,978	1,003	1,000	1,029	1,003
84 m	0,553	0,975	1,004	0,963	0,972	1,004
96 m	0,557	0,976	1,004	0,963	0,986	1,005
120 m	0,560	0,987	1,004	0,966	0,984	1,005

Fonte: Elaboração própria.

Nos demais prazos de projeção, de 3, 6 e 12 meses, conforme mostram as tabelas 7, 8 e 9 respectivamente, nenhum modelo apresentou resultado superior ao Passeio Aleatório de forma sistemática, da mesma forma que o exercício utilizando somente os dados das curvas de juros. Para esses horizontes, os modelos só foram superiores ao PA na parte mais curta da curva, sendo que, à medida que o prazo de projeção se torna mais longo, mais o Passeio Aleatório se torna superior aos demais modelos.

Para o prazo de projeções de 3 meses, conforme apresentado na Tabela 7, os modelos foram superiores ao PA somente nos vértices de curto prazo, principalmente até o vértice de 18 meses, sendo que os modelos LASSO e Floresta Aleatória apresentaram o melhor desempenho até os vértices de 18 e 24 meses, respectivamente. Nesses dois modelos metade dos vértices apresentaram resultado superior ao PA. Já os modelos kNN e XGB só foram superiores até o vértice de 9 meses, enquanto o modelo SVM foi inferior ao PA em quase todos os vértices.

Já para o prazo de projeção de 6 meses, conforme Tabela 8, os modelos só foram superiores ao PA na parte curta da curva, até o vértice de 12 meses, nos modelos LASSO e Floresta Aleatória, sendo que este último ainda apresentou desempenho superior no vértice de 96 meses. Já o modelo SVM só apresentou resultado melhor que o PA para os vértices de 9 e 12 meses.

Com relação ao prazo de 12 meses, apresentado na Tabela 9, o PA foi superior aos modelos na maioria dos vértices.

Tabela 7 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 3 meses

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	0,614	0,742	0,992	0,584	0,577	0,972
2 m	0,612	0,749	0,992	0,625	0,640	1,232
3 m	0,613	0,791	0,992	0,650	0,708	1,440
4 m	0,618	0,839	0,992	0,674	0,734	1,468
6 m	0,629	0,903	0,992	0,767	0,894	1,392
9 m	0,661	0,969	0,994	0,850	0,961	0,991
12 m	0,688	1,015	0,996	0,938	1,038	0,992
18 m	0,745	1,054	0,999	1,008	1,099	0,997
24 m	0,779	1,062	1,001	0,991	1,103	1,000
30 m	0,816	1,063	1,003	1,043	1,116	1,002
36 m	0,830	1,066	1,010	1,031	1,093	1,008
48 m	0,871	1,054	1,006	1,034	1,063	1,006
60 m	0,892	1,051	1,007	1,029	1,067	1,007
84 m	0,916	1,044	1,008	1,017	1,019	1,010
96 m	0,923	1,046	1,009	1,005	1,056	1,011
120 m	0,932	1,044	1,010	1,014	1,028	1,012

Fonte: Elaboração própria.

Tabela 8 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 6 meses

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	1,074	0,894	0,987	0,678	0,716	1,162
2 m	1,061	0,893	0,996	0,727	0,773	1,335
3 m	1,061	0,934	0,996	0,752	0,833	1,517
4 m	1,064	0,980	0,996	0,781	0,874	1,542
6 m	1,069	1,021	0,996	0,867	0,971	1,486
9 m	1,086	1,061	0,996	0,926	1,047	0,993
12 m	1,099	1,097	0,997	0,994	1,135	0,993
18 m	1,122	1,135	1,009	1,044	1,150	1,006
24 m	1,153	1,126	1,003	1,027	1,149	1,001
30 m	1,189	1,121	1,007	1,060	1,102	1,004
36 m	1,205	1,116	1,009	1,041	1,080	1,006
48 m	1,241	1,111	1,012	1,030	1,070	1,011
60 m	1,265	1,106	1,014	1,050	1,087	1,013
84 m	1,293	1,098	1,016	1,031	1,079	1,018
96 m	1,306	1,101	1,017	0,994	1,095	1,020
120 m	1,319	1,098	1,018	1,047	1,067	1,021

Fonte: Elaboração própria.

Tabela 9 – Resultados Curvas de Juros e Variáveis Macroeconômicas, horizonte de 12 meses

Prazo	REQM	TREQM				
	PA	kNN	LASSO	RF	XGB	SVM
1 m	1,895	1,133	1,006	0,880	0,906	2,073
2 m	1,891	1,094	1,005	0,907	0,958	1,962
3 m	1,886	1,120	1,004	0,927	1,035	1,990
4 m	1,883	1,146	1,004	0,944	1,056	2,026
6 m	1,871	1,144	1,003	1,010	1,108	1,842
9 m	1,847	1,149	1,001	1,036	1,206	0,998
12 m	1,814	1,166	0,999	1,074	1,256	0,994
18 m	1,753	1,173	0,998	1,076	1,198	0,994
24 m	1,720	1,167	1,000	1,066	1,188	0,996
30 m	1,715	1,161	1,004	1,069	1,140	0,999
36 m	1,711	1,156	1,008	1,057	1,119	1,002
48 m	1,716	1,153	1,014	1,037	1,118	1,012
60 m	1,723	1,151	1,017	1,074	1,135	1,015
84 m	1,750	1,145	1,022	1,047	1,136	1,026
96 m	1,766	1,145	1,024	0,986	1,156	1,029
120 m	1,780	1,144	1,027	1,057	1,098	1,032

Fonte: Elaboração própria.

Por fim, a inclusão das variáveis macroeconômicas para o treinamento dos modelos não trouxe ganhos de forma sistemática à capacidade de projeção, apesar da existência de correlação entre essas variáveis e diferentes vértices das curvas de juros.

Na Tabela 10 a seguir, são apresentados os resultados do TReREQM entre os resultados dos modelos treinados utilizando as variáveis macroeconômicas em conjunto com os dados das curvas de juros e dos modelos treinados somente com os dados históricos das curvas de juros.

Valores inferiores a 1 indicam que a inclusão das variáveis macro nos modelos gerou resultados superiores aos modelos treinados somente com os dados das curvas de juros. Para o modelo LASSO, em particular, com exceção do primeiro vértice, os resultados foram os mesmos nas duas situações. Isso ocorre porque os coeficientes que são irrelevantes para este modelo são zerados, portanto este modelo, nas duas situações, utilizam as mesmas covariadas para as projeções.

Tabela 10 – TReREQM entre os modelos com diferentes covariadas

TReREQM																
Projeção para 12 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	1,022	0,972	0,993	1,004	1,009	1,002	1,001	1,002	1,010	1,017	1,022	1,029	1,033	1,036	1,039	1,042
LASSO	1,001	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RF	0,960	0,975	0,973	0,982	0,979	0,959	0,968	0,972	0,950	0,980	0,956	0,967	0,999	0,981	0,931	0,988
XGB	0,887	0,892	0,924	0,914	0,884	0,933	1,048	0,957	1,003	0,989	0,975	0,992	1,070	0,996	1,075	0,984
SVM	2,195	2,033	1,953	1,991	1,682	0,987	0,995	0,991	1,000	0,995	0,997	0,997	0,999	1,002	0,999	0,997
Projeção para 6 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	1,022	0,973	0,990	1,003	1,005	1,002	1,001	1,002	1,007	1,010	1,012	1,016	1,017	1,017	1,023	1,026
LASSO	1,004	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RF	0,961	0,993	0,997	0,990	0,993	0,986	0,996	1,009	0,974	1,019	0,979	0,976	0,982	0,986	0,948	0,995
XGB	0,980	0,954	0,960	0,971	0,913	0,937	1,040	1,017	1,004	0,977	0,961	0,980	0,989	0,974	1,022	0,976
SVM	1,562	1,682	1,797	1,826	1,472	0,999	1,001	0,995	1,000	0,999	1,001	0,999	1,000	1,001	0,999	0,998
Projeção para 3 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	1,002	0,963	0,972	0,982	0,984	0,991	0,996	1,000	1,004	1,007	1,009	1,012	1,012	1,012	1,017	1,021
LASSO	1,005	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RF	0,981	0,998	1,024	0,985	0,996	0,997	1,022	1,030	0,993	1,027	1,008	1,002	0,977	0,990	0,976	0,989
XGB	1,001	0,982	0,993	0,955	0,981	0,924	1,039	1,022	0,995	0,982	0,941	0,993	0,968	0,956	0,979	0,982
SVM	1,512	1,730	1,835	1,891	1,397	0,999	1,004	0,997	1,000	1,000	1,000	1,000	1,000	1,000	0,999	1,000
Projeção para 1 mês																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,993	0,959	0,959	0,965	0,971	0,984	0,991	0,998	1,002	1,005	1,006	1,008	1,006	1,005	1,008	1,021
LASSO	1,008	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RF	0,998	0,994	1,027	1,052	0,989	0,971	0,992	1,012	1,007	1,037	1,077	1,025	1,007	0,970	0,973	0,982
XGB	0,927	1,002	0,982	0,949	0,934	0,970	1,036	0,930	1,068	1,011	0,992	1,002	1,031	0,956	0,931	0,921
SVM	1,494	1,737	1,807	1,755	1,420	0,993	1,003	0,999	1,000	1,002	1,004	0,997	0,994	1,002	0,995	1,000

Fonte: Elaboração própria.

Com relação aos testes de Diebold-Mariano – DM, os resultados mostram que, tanto para as projeções usando somente os dados das curvas de juros, quanto para as projeções utilizando também as variáveis macroeconômicas, com algumas poucas exceções, a hipótese nula somente é rejeitada na parte curta da curva, até o vértice de 12 meses, considerando p-valores inferiores a 10%. As Tabelas 11 e 12, apresentam o p-valores dos testes.

A hipótese alternativa confrontada com a hipótese nula é a de que os modelos de aprendizado de máquina apresentariam resultados estatisticamente superiores ao do Passeio Aleatório.

Tabela 11 – Teste de DM – Covariadas: Curvas de Juros

p-valor																
Projeção para 12 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,874	0,912	0,930	0,953	0,959	0,986	0,996	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,998
LASSO	0,641	0,660	0,641	0,626	0,591	0,536	0,482	0,453	0,501	0,584	0,664	0,781	0,851	0,926	0,942	0,958
RF	0,126	0,172	0,251	0,282	0,692	0,902	0,951	0,964	0,990	0,975	0,989	0,983	0,977	0,958	0,966	0,989
XGB	0,573	0,756	0,868	0,910	0,988	0,988	0,988	0,990	0,998	0,982	0,993	0,998	0,934	0,997	0,960	0,995
SVM	0,229	0,298	0,609	0,603	0,929	0,778	0,475	0,581	0,402	0,603	0,604	0,799	0,838	0,916	0,953	0,973
Projeção para 6 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,116	0,191	0,262	0,396	0,579	0,811	0,944	0,991	0,994	0,995	0,996	0,996	0,997	0,997	0,996	0,994
LASSO	0,242	0,353	0,342	0,342	0,345	0,378	0,410	0,507	0,583	0,670	0,723	0,810	0,859	0,907	0,919	0,929
RF	0,003	0,004	0,005	0,010	0,046	0,177	0,484	0,711	0,908	0,868	0,962	0,943	0,975	0,889	0,951	0,982
XGB	0,009	0,033	0,068	0,161	0,762	0,890	0,896	0,960	0,996	0,998	0,997	0,996	0,996	0,997	0,952	0,982
SVM	0,008	0,014	0,039	0,024	0,575	0,327	0,270	0,569	0,528	0,669	0,644	0,797	0,844	0,894	0,926	0,949
Projeção para 3 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,004	0,010	0,020	0,049	0,154	0,372	0,619	0,858	0,904	0,924	0,922	0,901	0,897	0,861	0,843	0,781
LASSO	0,072	0,170	0,171	0,183	0,212	0,290	0,349	0,471	0,541	0,612	0,655	0,730	0,776	0,824	0,836	0,842
RF	0,000	0,001	0,001	0,001	0,004	0,026	0,084	0,307	0,486	0,659	0,688	0,786	0,921	0,780	0,818	0,774
XGB	0,000	0,001	0,002	0,018	0,162	0,666	0,490	0,930	0,946	0,998	0,996	0,953	0,984	0,946	0,979	0,861
SVM	0,002	0,006	0,025	0,010	0,472	0,217	0,152	0,513	0,509	0,567	0,628	0,704	0,748	0,826	0,854	0,858
Projeção para 1 mês																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,000	0,000	0,001	0,003	0,020	0,069	0,154	0,223	0,215	0,204	0,174	0,129	0,130	0,109	0,105	0,096
LASSO	0,024	0,059	0,059	0,080	0,121	0,222	0,294	0,417	0,480	0,537	0,577	0,638	0,682	0,724	0,734	0,739
RF	0,000	0,000	0,000	0,000	0,002	0,011	0,070	0,106	0,140	0,108	0,052	0,375	0,410	0,406	0,378	0,280
XGB	0,000	0,000	0,001	0,006	0,054	0,274	0,152	0,735	0,317	0,827	0,731	0,491	0,486	0,631	0,926	0,839
SVM	0,000	0,001	0,005	0,011	0,815	0,408	0,083	0,408	0,420	0,341	0,318	0,704	0,826	0,642	0,867	0,748

Fonte: Elaboração própria.

Tabela 12 – Teste de DM – Covariadas: Curvas de Juros e Variáveis Macroeconômicas

p-valor																
Projeção para 12 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,889	0,829	0,893	0,936	0,954	0,981	0,994	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
LASSO	0,679	0,660	0,641	0,626	0,591	0,536	0,482	0,453	0,501	0,584	0,664	0,781	0,851	0,926	0,942	0,958
RF	0,038	0,089	0,137	0,193	0,566	0,737	0,904	0,931	0,942	0,966	0,916	0,833	0,967	0,928	0,362	0,970
XGB	0,137	0,319	0,658	0,775	0,894	0,996	0,998	0,986	0,980	0,998	0,999	0,997	0,995	0,997	0,997	0,987
SVM	0,990	0,999	0,999	0,998	0,986	0,439	0,316	0,345	0,400	0,477	0,540	0,756	0,824	0,935	0,953	0,969
Projeção para 6 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,186	0,149	0,253	0,418	0,592	0,792	0,926	0,987	0,994	0,997	0,997	0,998	0,999	0,998	0,998	0,997
LASSO	0,370	0,353	0,342	0,342	0,345	0,378	0,410	0,507	0,583	0,670	0,723	0,810	0,859	0,907	0,919	0,929
RF	0,001	0,003	0,005	0,007	0,034	0,111	0,452	0,751	0,783	0,933	0,786	0,774	0,919	0,830	0,430	0,952
XGB	0,007	0,015	0,049	0,080	0,377	0,737	0,980	0,976	0,958	0,989	0,984	0,993	0,986	0,981	0,989	0,973
SVM	0,891	1,000	0,992	0,987	0,975	0,274	0,279	0,437	0,534	0,639	0,681	0,796	0,844	0,912	0,928	0,938
Projeção para 3 meses																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,008	0,009	0,021	0,051	0,142	0,342	0,589	0,843	0,913	0,942	0,949	0,945	0,942	0,919	0,917	0,902
LASSO	0,177	0,170	0,171	0,183	0,212	0,290	0,349	0,471	0,541	0,612	0,655	0,730	0,776	0,824	0,836	0,842
RF	0,000	0,000	0,001	0,001	0,003	0,017	0,160	0,546	0,399	0,835	0,670	0,755	0,766	0,668	0,549	0,681
XGB	0,000	0,001	0,004	0,005	0,144	0,326	0,737	0,946	0,922	0,974	0,992	0,949	0,953	0,699	0,966	0,828
SVM	0,421	0,996	0,993	0,988	0,980	0,182	0,238	0,412	0,508	0,588	0,623	0,715	0,755	0,831	0,848	0,857
Projeção para 1 mês																
	1 m	2 m	3 m	4 m	6 m	9 m	12 m	18 m	24 m	30 m	36 m	48 m	60 m	84 m	96 m	120 m
kNN	0,001	0,000	0,001	0,002	0,014	0,054	0,128	0,204	0,234	0,257	0,243	0,210	0,203	0,164	0,188	0,333
LASSO	0,075	0,059	0,059	0,080	0,121	0,222	0,294	0,417	0,480	0,537	0,577	0,638	0,682	0,724	0,734	0,739
RF	0,000	0,000	0,000	0,000	0,001	0,002	0,048	0,184	0,161	0,367	0,648	0,643	0,503	0,147	0,157	0,088
XGB	0,000	0,000	0,001	0,001	0,015	0,170	0,258	0,347	0,717	0,896	0,759	0,507	0,740	0,298	0,402	0,308
SVM	0,141	0,978	0,997	0,996	0,986	0,105	0,153	0,356	0,443	0,444	0,524	0,610	0,669	0,736	0,751	0,762

Fonte: Elaboração própria.

Nota-se que, quanto mais curto o horizonte de projeções, melhores são os resultados dos modelos de aprendizado de máquina com relação ao Passeio Aleatório. Contudo, a hipótese nula não é rejeitada na maioria dos vértices, indicando que os modelos apresentam, nesse caso, projeções estatisticamente semelhantes entre os modelos testados, mesmo para o horizonte de projeção de 1 mês, ou seja, não pode-se afirmar que os modelos sejam superiores ao PA.

7. Conclusão

Este trabalho buscou analisar diferentes modelos de aprendizado de máquina que possibilitassem obter projeções da ETTJ do Brasil superiores ao Passeio Aleatório – PA, usado na literatura como referência para avaliação de desempenho de modelos de projeções de séries temporais. Para calcular a ETTJ foram utilizadas as taxas dos títulos prefixados (LTN e NTN-F) no mercado secundário, obtidas na Associação Brasileira das Entidades dos Mercados Financeiros e de Capitais – ANBIMA.

Com base nos modelos de aprendizado de máquina analisados foram obtidas as projeções da ETTJ para os horizontes de 1, 3, 6 e 12 meses. Os resultados obtidos levam à conclusão que não é possível determinar um modelo, para os dados brasileiros, que consistentemente produza menor erro de previsão que o PA, de forma similar ao obtido por Araujo (2011) e Cajueiro *et al.* (2009). Contudo, os modelos superaram o PA nos vértices mais curtos da curva de juros, principalmente nas projeções com horizonte até 6 meses. Adicionalmente, quanto mais curto o prazo de projeção, menores são os erros obtidos. Para o prazo de projeção de 1 mês, tanto o modelo kNN quanto o modelo Floresta Aleatória foram superiores ao PA em todos os vértices. Já os testes de Diebold-Mariano apontaram que, somente na parte curta da curva, até o vértice de 12 meses, as projeções dos modelos de Aprendizado de Máquina se mostraram estatisticamente superiores ao PA.

As análises mostram que o resultado de cada modelo depende do horizonte de projeções, do prazo de maturidade de cada vértice da ETTJ e do período do tempo em que foram feitas as projeções. Tal resultado indica que diferentes modelos de aprendizado de máquina podem se complementar para efetuar projeções das curvas de juros com menor erro que modelos individuais.

Os resultados também indicam que, nos modelos de aprendizado de máquina utilizados neste trabalho, a inclusão de variáveis macroeconômicas para o treinamento não

gerou melhores resultados nas projeções de forma consistente, apesar da existência de correlação entre essas variáveis e a dinâmica das taxas de juros da ETTJ.

Como sugestão de trabalhos futuros, é interessante que sejam analisadas combinações de diferentes modelos de aprendizado de máquina de modo a avaliar se essas combinações geram resultados superiores aos modelos individuais. Também seria válido trabalhar com dados diários e avaliar a capacidade de projeção dos modelos para prazos inferiores há um mês.

8. Referências Bibliográficas

ANG, A.; PIAZZESI, M. **A no-arbitrage Vector Autoregression of Term Structure dynamics with macroeconomic and latent variables**. Journal of Monetary Economics, v. 50, p. 745-787, 2003.

ARAUJO, R. **Combinações de Modelos de Previsão da Estrutura a Termo da Taxa de Juros: Aplicações ao Caso Brasileiro**, 2011.

BERGSTRA, J.; BENGIO, Y. **Random Search for Hyper-Parameter Optimization**. Journal of Machine Learning Research, 2012.

BLACK, F.; DERMAN, E.; TOY, W. **A one-factor model of interest rates and its application to Treasury bond options**. Financial Analysts Journal, p. 33-39, Jan-Fev, 1990.

BRENNAN, M.; SCHWARTZ, E. **A continuous time approach to the pricing bonds**. Journal of Banking and Finance, v.3, p. 133-155, 1979.

BOWSER, C. G.; MEEKS, R. **The Dynamics of Economic Functions: Modelling and Forecasting the Yield Curve**. Journal of the American Statistical Association. Mar 2008

- BOSER, B.E.; GUYON, I.M.; VAPNIK, V.N. **A Training Algorithm for Optimal Margin Classifiers**. Annual Workshop on Computational Learning Theory (COLT'92), Pittsburgh, 27-29 Julho, 1992, p 144-152.
- CAJUEIRO, D. O.; DIVINO, J. A.; TABAK, B.M. **Forecasting the yield curve for Brazil**. Working paper 197, Banco Central do Brasil, nov, 2009.
- COX, J. C.; INGERSOLL, J.E.; ROSS, S.A. **A Theory of the term structure of interest rates**. *Econometrica*, v.53, p. 385-407, 1985.
- CHRISTOFFERSEN, P. F.; DIEBOLD, F. X. **Cointegration and Long-Horizon Forecasting**. *Journal of Business & Economic Statistics*, v. 16, p. 450-458, 1998.
- DIEBOLD, F. X.; LI, C. **Forecasting the term structure of government bond yields**. *Journal of Econometrics*, Elsevier, v. 130, n. 2, p. 337-364, 2006.
- DIEBOLD, F. X.; LI, C.; YUE, V. Z. **Global yield curve dynamics and interactions: a dynamic Nelson-Siegel approach**. *Journal of Econometrics*, Elsevier, v. 146, n. 2, p. 351-363, 2008.
- DIEBOLD, F. X.; MARIANO, R. S. **Comparing Predictive Accuracy**. *Journal of Business and Economic Statistics*, 13, 253-263, 1995.
- DIEBOLD, F. X. **Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests**. *Journal of Business and Economic Statistics*, 2013.
- DUFFEE, G. R. **Term premia and interest rate forecasts in affine models**. *The Journal of Finance*, Wiley Online Library, v. 57, n. 1, p. 405-443, 2002.
- GUILLEN, O. T.; TABAK, B.M. **Characterizing the Brazilian Term Structure of interest rates**. *International Journal of Monetary Economics and Finance*, v. 2, no. 2, p. 103-114. Jan 2009.

- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Second Edition. Springer. 2017.
- HEATH, D.; JARROW, R.; MORTON, A. **Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation**. *Econometrica*, v. 60, p. 77-105, 1992.
- HO, S.Y.; LEE S-B. **Term Structure movements and pricing interest rate contingent claims**. *Journal of Finance*, v. 41, p. 1011-1029, 1986.
- HO, T. K. **Random Decision Forest**. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 14-16, 278-282, August 1995.
- HORDAL, P.; TRISTANI, O.; VESTIN, D. **A joint Econometric Model of Macroeconomic and Term-Structure Dynamics**. *Journal of Econometrics*, v. 131, p. 405-444, 2006.
- HULL, J.; WHITE, A. **Pricing interest rate derivative securities**. *Review Financial Studies*, v.3, p. 573-592, 1990
- KOOPMAN, S. J.; MALLEE, M. I.; WEL, M. Van der. **Analyzing the term structure of interest rates using the dynamic Nelson{Siegel model with time-varying parameters**. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 28, n. 3, p. 329-343, 2010.
- LEITE. A. L.; GOMES FILHO. R.; VICENTE, J. **Previsão da curva de juros: um modelo estatístico com variáveis macroeconômicas**. Working paper 186, Banco Central do Brasil, mai, 2009.
- LONGSTAFF, F. A.; SCHARTZ, E.S. **Interest Rate Volatility and the Term Structure: A two-factor general equilibrium model**. *Journal of Finance*, v.47, p. 1259-1282, 1992.
- LUDVINGSON, S.C.; NG, S., **Macro Factors in bond risk premia**. *Review of Financial Studies*, 2009.

- MONCH, E. **Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach.** Working paper 544, European Central Bank, 2008.
- NELSON, C. R.; SIEGEL, A. F. **Parsimonious modeling of yield curves.** The Journal of Business, v. 60, p. 473-489, out, 1987.
- POOTER, M. D. **Examining the Nelson-Siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance.** Available at SSRN 992748, 2007.
- POOTER, M. D.; RAVAZZOLO, F.; DIJK, D. J. V. **Term structure forecasting using macro factors and forecast combination.** FRB International Finance Discussion Paper, n. 993, 2010.
- SARADHI, V.; KAMIK, H.; MITRA, P. **A Decomposition Method for Support Vector Clustering.** In Proc. of the 2nd International Conference on Intelligent Sensing and Information Processing (ICISIP), p. 268-271, 2005.
- SVENSSON, L. E. O. **Estimating forward interest rates with the extended Nelson-Siegel Method.** Quarterly Review, v. 3, p. 13-26, Sveriges Riskbank, 1995.
- VAPINIK, V. N.; GOLOWICH, S.; SMOLA, A. J. **Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing.** Advances in Neural Information Processing Systems 9, p. 281-287. 1997.
- VASICEK, O. **An equilibrium characterization of the Term Structure.** Journal of Financial Economics, v.5, p.177-188, 1977.