



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE ECONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Flávio Garrido Torres

PREVENDO A INFLAÇÃO BRASILEIRA: ESTIMANDO O IPCA POR DIFERENTES  
MÉTRICAS

Rio de Janeiro

2025

Flávio Garrido Torres

PREVENDO A INFLAÇÃO BRASILEIRA: ESTIMANDO O IPCA POR DIFERENTES  
MÉTRICAS

Dissertação apresentada ao Programa de Pós-  
Graduação em Economia da Universidade  
Federal do Rio de Janeiro, como requisito para  
a obtenção do título de Mestre em Economia.

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Viviane Luporini

Rio de Janeiro  
2025

## FICHA CATALOGRÁFICA

T693p Torres, Flávio Garrido.  
Previendo a inflação brasileira: estimando o IPCA por diferentes métricas /  
Flávio Garrido Torres. – 2025.  
74 f.

Orientadora: Viviane Patrizzi Luporini.  
Dissertação (mestrado) – Universidade Federal do Rio de Janeiro, Instituto de  
Economia, Programa de Pós-Graduação em Economia da Indústria e da Tecnologia,  
2025.  
Bibliografia: f. 71 – 74.

1. Inflação. 2. *Machine learning*. 3. Séries temporais. I. Luporini, Viviane  
Patrizzi, orient. II. Universidade Federal do Rio de Janeiro. Instituto de Economia.  
III. Título.

CDD 332.4

Flávio Garrido Torres

PREVENDO A INFLAÇÃO BRASILEIRA: ESTIMANDO O IPCA POR DIFERENTES  
MÉTRICAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Rio de Janeiro, como requisito para a obtenção do título de Mestre em Economia.

Rio de Janeiro, 19 de fevereiro de 2025.

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Viviane Luporini - Orientadora  
Universidade Federal do Rio de Janeiro (UFRJ)

---

Prof. Dr. Pedro Hemsley – Membro Interno  
Universidade Federal do Rio de Janeiro (UFRJ)

---

Prof. Dr. Luciano Vereda – Membro Externo  
Universidade Federal Fluminense (UFF)

**Aos meus pais,  
que sempre me apoiaram.**

## **AGRADECIMENTOS**

A finalização desta dissertação só foi possível por conta da ajuda de muitas pessoas ao longo da minha vida pessoal e acadêmica, especialmente nesses últimos anos. A minha gratidão por todas elas será eterna.

Em primeiro lugar, agradeço à minha mamãe, Maria José Clarisse Garrido Torres, que me apoiou de forma incondicional durante o tempo em que estive aqui, sempre com muito amor, carinho e bondade. Se chego neste momento da minha vida acadêmica, é quase que exclusivamente em virtude dela, já que, em muitos momentos, abriu mão de diversas incumbências para que eu seguisse. Deste modo, agradeço e dedico esta dissertação em sua memória.

Não menos importante, cabe dedicação especial e agradecimento ao meu papai, Lealdino Teixeira Torres, que sempre me aconselhou em toda a minha vida, e, embora não entendesse tão bem o caminho que eu estava trilhando, respaldava-me incessantemente. Também dedico esta dissertação.

Por fim, agradeço à excelente professora e minha orientadora, Viviane Luporini, por toda paciência nesse período, além das diversas sugestões e atenção. Não foi um trabalho exatamente fácil de se concluir, mas, por intermédio de sua sabedoria e sagacidade, tornou-se muito mais factível. Obrigado.

## RESUMO

Prever a inflação é uma tarefa difícil, mas que é muito necessária para o dia a dia dos agentes econômicos. Se eles tiverem alguma noção da tendência do nível de preços em um certo horizonte, poderão tomar decisões que os beneficiem em um prazo determinado. Por conta dessa demanda dos indivíduos de um país, ocorre que diversas instituições e pesquisadores independentes procuram desenvolver e estimar a inflação por variados conjuntos de técnicas. Assim, o objetivo desta dissertação é fazer a previsão do índice de preços ao consumidor brasileiro (IPCA) usando tanto técnicas mais tradicionais, como as de séries temporais, mas também algumas que são elencadas como mais recentes, que é o caso das de machine learning. Uma vez feito esse exercício, serão comparados os erros de previsão dos modelos entre si para constatar qual ou quais deles apresentam os melhores resultados. Além disso, serão realizados testes para saber se os erros de previsão são estatisticamente significativos, ou seja, se a acurácia de um método é realmente superior à outra. Em termos gerais, concluiu-se que os métodos de machine learning têm erros de previsão menores do que as técnicas de séries temporais para quase todos os horizontes, sendo o Support Vector Machine aquele que melhor performou. Contudo, quando se observou minuciosamente a significância desses erros, percebeu-se que todos os modelos são iguais estatisticamente.

**Palavras-chave:** Previsão de inflação; Machine learning; Séries temporais.

## ABSTRACT

Forecasting inflation is a difficult task, but pretty important on a daily basis for economic agents. If they somehow know where the price level trend is going in a certain horizon, they will be able to make good decisions in a short to medium term. Because of the demand from individuals in this country, many institutions and independent researchers try to develop and estimate inflation using different techniques. The aim of this study is to forecast the Brazilian Consumer Price Index (IPCA) using time series methods as well as machine learning models. Once this is done, a comparison will be made among the methods in terms of forecasting errors to know which one has the smallest Mean Squared Error and Root Mean Squared Error. Additionally, another test will be implemented to assess whether the forecast errors are statistically significant, i.e., to determine if one predictive accuracy is indeed superior to another. Generally speaking, the results showed that machine learning methods have smaller forecast errors relative to time series models for almost all horizons, and the Support Vector Machine was the best performer among all. Regarding error significance, however, it was found that all the models are statistically equivalent.

**Keywords:** Inflation Forecasting; Machine Learning; Time Series.

## SUMÁRIO

|   |    |
|---|----|
| <b>INTRODUÇÃO</b> .....   | 11 |
| <b>CAPÍTULO I – CARACTERÍSTICAS DA INFLAÇÃO E PREVISÃO</b> .....                    | 14 |
| 1.1 Revisão de literatura: os modelos de previsão de inflação mais frequentes ..... | 14 |
| 1.2 Alguns aspectos teóricos da inflação .....                                      | 16 |
| 1.3 Núcleo da inflação .....  | 19 |
| 1.4 A importância da previsão .....   | 23 |
| 1.5 Alguns potenciais problemas ao fazer previsão .....                             | 27 |
| <b>CAPÍTULO II – METODOLOGIA</b> .....  | 30 |
| 2.1 Base de dados .....   | 30 |
| 2.2 Modelos .....   | 33 |
| 2.3 Critérios de avaliação .....  | 46 |
| <b>CAPÍTULO III – RESULTADOS</b> .....  | 50 |
| 3.1 Modelos .....   | 50 |
| 3.2 Erros de previsão .....   | 59 |
| 3.3 Teste de Diebold-Mariano .....  | 61 |
| <b>CONCLUSÃO</b> .....  | 67 |
| <b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....   | 69 |

## LISTA DE FIGURAS

|   |  |
|---|--|
| Figura 1 - Função de autocorrelação – IPCA.....               |  |
| Figura 2 - Função de autocorrelação parcial – IPCA.....       |  |
| Figura 3 - Previsão de KNN para $k = 3$ .....                 |  |
| Figura 4 - Número de k-vizinhos escolhidos no software.....   |  |
| Figura 5 - Representação de um SVM linearmente separável..... |  |
| Figura 6 - AR x IPCA.....                                     |  |
| Figura 7 - ARMA x IPCA.....                                   |  |
| Figura 8 - KNN x IPCA.....                                    |  |
| Figura 9 - SVM x IPCA.....                                    |  |
| Figura 10 - Light GBM x IPCA.....                             |  |
| Figura 11 - Métodos x IPCA.....                               |  |

## **LISTA DE TABELAS**

Tabela 1 - Variáveis usadas e suas informações

Tabela 2 - Erros de previsão modelo AR

Tabela 3 – Erros de previsão modelo KNN

Tabela 4 - Erros de previsão modelo SVM

Tabela 5 - Erros de previsão modelo Light GBM

Tabela 6 - Erros de previsão dos principais modelos

Tabela 7 - Teste de Diebold-Mariano - AR x outros modelos

Tabela 8 - Teste de Diebold-Mariano - ARMA x outros modelos

Tabela 9 - Teste de Diebold-Mariano - KNN x outros modelos

Tabela 10 - Teste de Diebold-Mariano - SVM x outros modelos

Tabela 11 - Teste de Diebold-Mariano - Light GBM x outros modelos

## LISTA DE SIGLAS

ADF – Augmented Dickey-Fuller

AR – Autorregressivo

ARMA – Autorregressivo de Médias Móveis

BCB – Banco Central do Brasil

CP – Curva de Phillips

FAC – Função de Autocorrelação

FACP – Função de Autocorrelação Parcial

FED – Federal Reserve

FOMC – Federal Open Market Committee

IPCA – Índice de Preços ao Consumidor Amplo

KNN – k-Nearest Neighbors

MSE – Mean Squared Error

NAIRU – Non-Accelerating Inflation Rate of Unemployment

RMSE – Root Mean Squared Error

SVM – Support Vector Machines

TQM – Teoria Quantitativa da Moeda

## INTRODUÇÃO

A inflação é uma das variáveis mais importantes e estudadas na macroeconomia. Dentre os mais diversos tópicos em que ela está incluída, a sua previsão certamente postula entre os mais recorrentes. Se, por um lado, a previsão é uma tarefa extremamente difícil por fatores endógenos e até mesmo exógenos ao ambiente econômico, como o alto grau de incerteza que existe; por outro, é uma atividade bastante imprescindível por conta das potenciais consequências perversas que a inflação gera para a sociedade.

No sentido em que ela retira poder de compra dos indivíduos, as famílias constituem o grupo mais prejudicado por possuírem menos capacidade de se defender desse fenômeno, especialmente os mais pobres. Desse modo, é essencial que os integrantes de uma determinada economia consigam prever minimamente o comportamento futuro do nível de preços. Por exemplo, no caso do setor privado, pelo fato de haver contratos e obrigações de longo prazo indexados ao nível de preços, é fundamental que exista alguma forma de se antecipar a essas variações, instigando instituições privadas a desenvolverem métricas próprias de previsão.

Já no caso dos Bancos Centrais, esse processo é ainda mais importante, visto que eles têm como principal objetivo manter a estabilidade de preços do país, com uma inflação controlada, assim como, em alguns casos, também precisam manter o PIB próximo a um nível pré-determinado (Elliot & Timmermann, 2013). Os relatórios de inflação (Inflation Reports, nos quais o BC faz uma análise conjuntural sobre a economia como um todo) divulgados trimestralmente mostram a solidez e empenho nas projeções dos Bancos Centrais, além da transparência e maior chance de entendimento para o público geral.

Outra medida de grande relevância para a autoridade monetária é a trajetória do nível preços, que é capturada pelo núcleo de inflação (medida que desconsidera choques temporários sobre o índice). Os índices de núcleo ajudam o Banco Central no sentido de separar os choques que implicam em uma persistência no nível de preços daqueles que são puramente transitórios, e isso contribuirá para a ancoragem das expectativas. Resumidamente, a maioria dos países trabalha com uma metodologia que exclui do índice de preços tradicional alguns componentes como alimentos e energia (BCB, 2000). Apesar de haver certo debate sobre uma eventual substituição do índice cheio por um algum núcleo como índice oficial da inflação, o fato é que as médias dos núcleos ajudam consideravelmente o Banco Central a analisar a conjuntura e possíveis cenários para o nível de preços para então criar hipóteses para suas projeções, ajudando nas escolhas de política monetária.

Um primeiro motivo que incentiva os estudos sobre inflação no caso brasileiro é de motivo estrutural. Durante décadas, o país enfrentou períodos de alta inflação e planos de estabilização que foram, em sua maioria, desastrosos. Ainda que tenha tido êxito em controlar a hiperinflação na década de 1990, não é incomum que países emergentes tenham problema em erradicar por completo distúrbios nos seus níveis de preços. Exatamente por isso, o tema inflacionário é rotineiro no Brasil.

Tradicionalmente, os exercícios de previsão de inflação eram feitos por meio das técnicas mais conhecidas na área de econometria de séries temporais (Einav & Levin, 2014). Na realidade, para a previsão da maioria das variáveis macroeconômicas, com grande frequência ainda se usam esses métodos, que já se mostraram eficientes em uma série de ocasiões, e são formidáveis para o propósito que os pesquisadores estão interessados em muitos casos. Dentre eles, estão técnicas como processos autorregressivos, de médias móveis, diferentes especificações de um VAR, entre outros. Porém, talvez a abordagem mais comum ao se pensar em prever inflação seja a Curva de Phillips, que também admite algumas mudanças de parâmetros, mas que nem sempre se encaixa na série de inflação de um país, como será visto posteriormente.

Soma-se a isso o crescente número de técnicas computacionais que vêm tendo um uso cada vez maior na análise de dados nos últimos anos, como as de *machine learning*. Mesmo que muitas das ferramentas de aprendizado de máquina não sejam necessariamente novidades, no sentido de que já existiam há anos, foi apenas recentemente que a revolução na ciência de dados aumentou a utilização desses algoritmos. Esse crescimento ocorreu por conta do alto nível de previsões gerados, com uma grande acurácia e a possibilidade de usar um vasto conjunto de dados (*big data*). Os campos de aplicação desses métodos são gigantes: desde detecção de anomalias em sistemas financeiros (como fraude no cartão de crédito) ao reconhecimento de imagem e voz, sistemas de recomendação de produtos e serviços, classificação de textos, entre centenas de outros casos. Na economia, a inclusão de *machine learning* parece representar uma mudança na análise de séries temporais, dado que já é usado por diversas instituições financeiras, além de ratificado e estimulado por economistas de grande renome, como é o caso de Varian (2014) e Athey & Imbens (2019).

Diante disso, é comum que surja o questionamento: em virtude dessa empolgação, esses novos métodos teriam a capacidade de apresentarem resultados melhores do que aqueles que são normalmente obtidos pelas técnicas mais tradicionais? No caso brasileiro, que a série de inflação não é tão estável como a de países desenvolvidos, essas ferramentas teriam bons resultados como é comumente visto em artigos internacionais?

Decorre daí que o objetivo deste trabalho é tentar fazer a previsão do principal índice de inflação do Brasil, que é o índice de preços ao consumidor amplo (IPCA). Para que se atinja esse objetivo, foram usadas tanto técnicas de séries temporais mais clássicas, como um processo autorregressivo e um processo autorregressivo de médias móveis; assim como um grupo de métodos de aprendizado de máquina, que englobam os algoritmos de K-Nearest Neighbours (KNN), Support Vector Machines (SVM) e Light GBM. Ademais, construiu-se uma base de dados com dezenas de variáveis explicativas que compreendem o período de 2006 até 2017. Uma vez feita a previsão para cada uma dessas técnicas, elas serão comparadas entre si para saber qual delas terá o menor erro, e testadas em pares para saber se essa diferença tem alguma significância estatística.

O artigo de Araújo & Gaglianone (2023) é a inspiração predominante para este trabalho na medida em que reúne diversas técnicas e as compara em uma análise refinada. Aqui, busca-se complementar o referido artigo.

A principal contribuição deste estudo é trazer um grupo de técnicas de *machine learning* que até então não foi tão usado no caso brasileiro para se prever inflação. Até este momento, grande parte dos trabalhos usa as mesmas técnicas sob a justificativa de serem as melhores. Porém, como se sabe, com a constante evolução computacional, é importante que sejam testados novos modelos para fins de investigação.

Além desta introdução e da conclusão, esta dissertação está dividida da seguinte forma: no capítulo um, será feita um apanhado teórico dos principais pontos sobre a inflação, uma revisão de literatura, e algumas definições. O segundo capítulo trará a metodologia usada na parte empírica e os critérios de avaliação. No terceiro capítulo serão mostrados os resultados para cada um dos modelos e uma breve discussão sobre qual teve o melhor desempenho.

## CAPÍTULO I – CARACTERÍSTICAS DA INFLAÇÃO E PREVISÃO

Neste primeiro capítulo, será feita uma abordagem de cunho mais teórico sobre alguns pontos centrais do fenômeno inflacionário e o exercício de previsão de variáveis macroeconômicas. De início, uma breve revisão de literatura sobre previsão com o uso de técnicas econométricas e de machine learning para inflação. Em seguida, os principais aspectos teóricos da inflação serão apresentados, de maneira que facilite o entendimento do tema. Então, discutiremos a importância da previsão de certas variáveis para os *policymakers*.

### 1.1 Revisão de literatura: os modelos de previsão de inflação mais frequentes

Inicialmente, ao se falar sobre previsão de inflação, pode-se dizer que, grosso modo, é uma temática antiga e com alguma recorrência caso se considere o uso de econometria por séries temporais e Curva de Phillips (CP). Em relação ao último, Atkeson e Ohanian (2001) analisam três conjuntos de Curva de Phillips NAIRU (que é aquela em que a taxa de desemprego se relaciona com variações na taxa de inflação, e que existe um nível para essa taxa de desemprego que não acelera a inflação, isto é, a NAIRU) para decidir se essa ferramenta tem alguma utilidade na previsão da inflação americana - o primeiro conjunto de previsão vem de manuais de macroeconomia, mostrando a relação negativa entre a taxa de desemprego e variações na taxa de inflação; já o segundo consiste em dois modelos de CP modificados por Stock & Watson (1999a); enquanto o terceiro conjunto é de um modelo próprio do Federal Reserve (FED). Na prática, eles comparam esses três conjuntos com um modelo ingênuo (*naive model*) para o período de um ano. Um modelo ingênuo (*naive model*) consiste em um método bastante simples usado como benchmark em estudos de previsão, mas não necessariamente inferiores; na verdade, costumam apresentar bons resultados em muitos trabalhos. A conclusão na qual os autores chegam é de que a CP NAIRU não foi uma boa técnica de previsão entre os anos 1985 e 2000 para os Estados Unidos, uma vez que não produziu resultados melhores do que o modelo ingênuo. Para eles, se um modelo que supostamente possui uma relação econômica não consegue ter respostas melhores do que um simples modelo atóxico (passeio aleatório), então o primeiro deve ser deixado de lado.

Entretanto, não se pode dizer que o uso da CP foi abandonado pelos analistas, dado que ainda é uma ferramenta imprescindível para os Bancos Centrais em suas estimativas. Segundo Arruda, Ferreira e Castelar (2011), alguns autores perceberam que, adicionando fatores não lineares na CP, a previsão melhora consideravelmente. Isso é válido inclusive para a série

brasileira, que apresentou bons resultados em modelos com efeitos limiaries (*threshold*). Esses modelos com efeito limiar acabam por superar aqueles modelos ingênuos apresentando um erro de previsão menor.

Sobre os modelos de séries temporais, diversos trabalhos já foram realizados para a previsão de inflação. Um exemplo recente é o artigo de Lack (2006), que tentava prever a inflação suíça por meio de alguns modelos do tipo VAR. O autor estimou diversos modelos em nível e em primeira diferença, selecionou os melhores com base no menor RMSE, e então fez uma combinação entre eles. De uma forma geral, essa combinação conseguia produzir resultados melhores do que o modelo ingênuo, ainda que o autor tenha percebido que um modelo mais enxuto gerasse um erro de previsão menor, ou seja, que existiam variáveis na base de dados original que, além de não ajudarem a prever a inflação com sucesso, acabavam por aumentar o erro de previsão e gerar resultados que não eram satisfatórios. Lack chegou nessa conclusão ao rodar o modelo diversas vezes, retirando uma variável em cada nova simulação, e percebeu que o modelo ficava cada vez melhor. Esse é um ponto importante dado que, em uma primeira impressão, pode-se pensar que, quanto mais completo for um modelo, melhor serão seus resultados. No entanto, a literatura já mostrou que modelos mais simples, como um processo autorregressivo, muitas vezes apresentam desfechos mais interessantes do que modelos mais rebuscados, como é verdade para o caso da previsão do PIB (Elliot & Timmermann, 2013).

Mais recentemente, técnicas computacionais de *machine learning* passaram a ser usadas para modelos de previsão. Para o caso brasileiro, um artigo atual de referência que traz importantes conclusões é de Araújo & Gaglianone (2023). Os autores utilizam dados que vão de janeiro de 2004 até agosto de 2021, com observações mensais (212, portanto), sendo que as primeiras 120 dessas observações são usadas para treinar o modelo, enquanto as últimas 92 são utilizadas para a previsão pseudo fora da amostra e, então, testam dezenas de métodos preditivos, utilizando modelos de *machine learning*, de séries temporais, de Curva de Phillips, e de combinações entre eles. Em meio às diversas conclusões que são apresentadas no artigo, uma delas tem importância fundamental sobre o horizonte temporal das previsões, ao mostrar que os métodos de machine learning, especialmente aqueles baseados em árvore, como o random forest e o XGBoost, costumam apresentar menores erros de previsão em períodos mais longos em relação ao modelo considerado como *benchmark* (neste caso, um modelo ARMA), mas que só em alguns casos são estatisticamente significantes. No entanto, ainda se tratando de Brasil, Medeiros, Vasconcelos & Freitas (2016) constatam que o modelo AR é um bom previsor do IPCA.

Para os dados americanos de inflação, Joulil e Iaousse (2023) apontam que a técnica de K-Nearest Neighbours (KNN) apresenta os melhores resultados entre alguns modelos para a previsão de curto prazo, com baixo MSE e RMSE. Ademais, Rodríguez-Vargas (2020) também expõe o bom resultado do KNN para a previsão da inflação da Costa Rica, ainda que para muitos horizontes o desempenho do modelo não seja estatisticamente significativo. Vale ressaltar, no entanto, que os próprios Joulil e Iaousse salientam para o perigo de generalização de uma técnica, e que nem sempre ela terá um desempenho satisfatório para séries de países diferentes. Seguindo por esse rumo, Ivascu (2023) analisa o caso da inflação da Romênia, fazendo uso de diversos modelos de machine learning que se provaram convincentes em outros países. Interessante notar que os modelos foram superiores para o caso univariado em um horizonte curto, mas, em um prazo maior, todos ficaram abaixo de um AR (1). E no caso multivariado, todos foram inferiores a um VAR para qualquer horizonte. Porém, este último autor não fez qualquer teste para identificar se as diferenças são estatisticamente significantes.

## 1.2 Alguns aspectos teóricos da inflação

Sucintamente, a inflação pode ser definida como um aumento no nível geral de preços. Em suma, esse nível de preços é medido por um índice de preços ao consumidor (mais comum) ou pelo deflator do PIB (Blanchard, 2007). Em um trabalho mais específico, Roger (1998) ressalta que existem motivos teóricos e práticos para que o índice de preços ao consumidor seja mais empregado do que o deflator do PIB, ou então do que um índice de preços ao produtor. Pelo lado teórico, ele argumenta que uma das principais preocupações da política monetária (PM) é o bem-estar da população, e para que se atinja esse objetivo de melhorar a qualidade de vida de um grupo de pessoas, faz mais sentido concentrar esforços em um índice que reflita o custo de vida desses indivíduos, isto é, um índice ao consumidor. Ademais, pela imprevisibilidade dos preços futuros, e pelo fato de que as pessoas têm uma menor capacidade de se proteger dessas flutuações nos preços (e estarem em maior número) do que as empresas, torna-se mais lógico que a PM tente estabilizar um índice de preços ao consumidor ao invés de um índice de preço ao produtor. Do mesmo modo, é por meio do controle à inflação que o Banco Central (BC) aumenta sua credibilidade frente ao público, sem contar que, de maneira geral, as expectativas de inflação estão mais atreladas àquelas dos preços ao consumidor.

Pelo lado prático, em muitos países, sabe-se que os índices de preços ao consumidor recebem um maior aporte de recursos por parte das agências estatísticas para serem construídos, se comparado aos demais. Isso ocasiona em uma medida que é melhor elaborada. Ainda

segundo Roger (1998), esse argumento de uma visão mais prática é o que muitos Bancos Centrais usam como última justificativa para adoção de um índice ao consumidor, além do pretexto sobre a credibilidade.

Assim, pragmaticamente, a interpretação do índice é importante, pois ele irá dizer que os agentes de uma economia perdem poder de compra a depender de quão mais alto está o nível de preços. Isso ocorre porque, na maioria das vezes, preços e salários não evoluem proporcionalmente (inflação pura), como é o caso das aposentadorias em alguns países, e a distorção de preços relativos causados por defasagens de preços administrados (Blanchard, 2007).

Suas causas podem variar, sendo que determinadas correntes de pensamento advogam por certos tipos de causadores. Por um lado, existem algumas vertentes da ortodoxia. Uma delas, mais recente, é a Teoria Fiscal do Nível de Preços. Segundo Resende (2017), o nível de preços é função do passivo financeiro do governo (estoque de moeda e dívida) e o valor presente esperado dos superávits primários. Portanto, o nível de preços depende dos juros, da política fiscal e da dívida pública, de forma dinâmica e interdependente. Uma corrente mais antiga acredita que a inflação é um fenômeno puramente monetário; na Teoria Quantitativa da Moeda (TQM), uma vez existindo um aumento exagerado na oferta de moeda, segue-se um aumento exagerado na demanda por moeda, e isso implica em um aumento no nível de preço. Como relembra Lavoie (2014), a TQM aparece em diferentes versões, mas sempre com a ideia de que a inflação decorre dessa dinâmica. Ainda segundo o autor, é comum encontrar na economia mainstream a tese de que a inflação acelera de acordo com alguma disparidade entre duas variáveis. Ele menciona o caso da versão wickselliana da TQM, na qual a taxa de juros de mercado e a taxa natural divergem. Neste caso, quando a taxa de juros de mercado é menor do que a taxa natural, isso gera um hiato no produto, que leva a uma aceleração da taxa de inflação, dado que ocorreu um aumento na oferta de moeda. A tese aceleracionista mais conhecida advém da Curva de Phillips NAIRU, uma vez que é a especificação mais usada na prática e também suas variações, como visto em Atkenson & Ohanian (2001). Como se sabe, a CP exhibe um trade-off de curto prazo entre inflação e desemprego, com uma relação negativa entre eles. Na versão NAIRU, existe uma taxa de desemprego que não acelera a inflação (a NAIRU), ou seja, se a taxa de desemprego ficar em um determinado patamar, a taxa de inflação ficará imóvel (Atkenson & Ohanian, 2001). Porém, a partir do momento que a taxa de desemprego se descolar da NAIRU, para baixo, a inflação acelera.

Por outro lado, na heterodoxia pós-keynesiana, um aumento na oferta de moeda não causa um aumento no nível de preços pelo fato de que essa oferta é endógena (Weintraub, 1978;

Wray, 2012; *apud* Lavoie, 2014). De acordo com essa corrente de pensamento, é o conflito distributivo de renda que dita a dinâmica inflacionária, em especial no que se chama espiral de preços-salários. Lavoie (2014) explica que aumentos no nível de preços podem aparecer tanto quando o salário nominal crescer mais rápido do que a produtividade do trabalho, assim como quando houver um aumento no mark-up, ou então quando houver um aumento de custo dos materiais importados usados como insumos no processo de produção em relação ao custo total do trabalho. Basicamente, a inflação pode ocorrer em uma situação em que o salário real desejado pelos sindicatos é diferente do salário real vigente (*wage inflation*), assim como nos casos em que o mark-up almejado pelas firmas é menor do que o que efetivamente é praticado (*price inflation*).

No que diz respeito à dinâmica inflacionária, Mishkin (2007) analisa alguns componentes da série americana para entender certas mudanças nos últimos anos. Embora os resultados não possam necessariamente ser generalizados para o caso brasileiro e outros países, é interessante perceber como uma série de fatos estilizados foram modificados nos últimos anos. O primeiro exemplo tem a ver com a persistência de choques na inflação. O autor decompõe esses choques em dois componentes: um de tendência e um temporário - seguindo o modelo de Stock e Watson (2007). O que se observava no passado era uma persistência muito grande do choque devido ao componente relativo de tendência ser maior do que o temporário. Contudo, ao passar dos anos, notou-se que a inflação ficou cada vez menos persistente, isto é, ela voltaria ao seu nível inicial mais rapidamente. Além dessa persistência, outros fatos estilizados também apresentaram mudanças sustentadas nas últimas décadas, como a inclinação da CP, e o fraco efeito de outros tipos de choques na inflação.

Mishkin realça que essa mudança na dinâmica tem a ver com as alterações nas expectativas de inflação, especialmente as de longo prazo. Ou seja, no passado houve um período em que as expectativas estavam desancoradas, implicando também em uma desancoragem da inflação de tendência (*trend inflation*) e, por conseguinte, a persistência da inflação era mais forte. Com o passar dos anos, verificou-se uma ancoragem dessas expectativas, que resultou em uma diminuição dessa persistência. Mas o ponto principal levantado pelo autor, e que pode ser verdade para outros países, é que a mudança das expectativas afeta de igual modo os outros fatos estilizados, e a forma de acontecer essa alteração é por meio da política monetária (PM).

Em outras palavras, pode-se dizer que quando o Banco Central (BC) atua de forma mais firme no combate à inflação, além de manter o nível de preços em um estágio aceitável, as expectativas de longo prazo tendem a ficar bem ancoradas. Nesse sentido, uma boa condução

da PM resulta em uma modificação dos determinantes da dinâmica inflacionária, já que os fatos estilizados são impactados. Todavia, o BC não deve ser permissivo com choques na inflação ao pensar que essas mudanças são perenes e estruturais. Isso quer dizer que, uma vez ocorrido o choque, as autoridades monetárias precisam agir rapidamente e com firmeza para manter a inflação em um nível mais baixo, permitindo que sua credibilidade seja mantida, e as expectativas continuem ancoradas. Isso realça a importância de se possuir uma boa política monetária para o controle inflacionário. E mesmo que eventualmente a dinâmica de preços seja diferente para o caso brasileiro, determinados pontos ainda podem ser usados e incorporados no caso nacional.

### 1.3 Núcleo da inflação

Oficialmente, segundo o Banco Central do Brasil, o núcleo da inflação pode ser definido como uma “medida que procura captar a tendência dos preços, desconsiderando distúrbios resultantes de choques temporários.” (BCB, 2000). Basicamente, ao desconsiderar movimentos voláteis sobre o índice cheio, fica mais intuitivo para a autoridade monetária perceber as condições de demanda da economia e saber de fato quais preços foram afetados de forma permanente. Assim, o núcleo é o componente da inflação no qual o Banco Central deve ter uma maior preocupação (Ruch et al., 2019).

Entrando em detalhes, Roger (1998) explica que existem duas linhas de pensamento para definir o conceito de núcleo da inflação: a primeira delas enxerga como sendo uma inflação persistente, e a outra como uma inflação generalizada.

Enveredando para o primeiro grupo, a ideia é que o processo inflacionário pode ser dividido em duas partes, sendo que um deles é visto com um componente persistente, que é incorporado às expectativas, e o outro é um componente transitório, com efeito passageiro e que dificilmente será antecipado pelos agentes. Então, essa primeira corrente estabelece o núcleo da inflação como sendo esse fator persistente, representando a tendência do nível de preços. Em termos mais formais, pode-se se apresentar o processo inflacionário da seguinte forma (Roger, 1998):

$$\Pi_t = \Pi_t^{LR} + f(X_{t-1}) + v_t \quad (1)$$

onde  $\Pi_t$  é a taxa de inflação no período ‘t’,  $\Pi_t^{LR}$  é a taxa de inflação de tendência no período ‘t’,  $X_{t-1}$  é um componente que mede pressões de demanda, e  $v_t$  representa um

componente transitório na inflação. Dessa forma, a partir desse ponto de vista, o núcleo de inflação é definido como sendo os dois primeiros termos da equação do lado direito, isto é, o componente transitório, que representam choques de oferta, não entra na caracterização (Quah e Vahey *apud* Roger 1998).

Essa abordagem para o núcleo de inflação pode acabar permitindo choques de demanda que tenham apenas efeitos cíclicos. Também existe na literatura uma forma mais restritiva de definir o núcleo de inflação pela corrente do componente permanente, só que desta vez, excluindo esses choques de demanda, fazendo então com que o núcleo seja apenas o primeiro termo do lado direito da equação 1 (Eckstein *apud* Roger 1998). No geral, a diferença só terá importância ao se escolher o horizonte temporal da política monetária; se for mais alinhada ao médio prazo, a primeira delas será a melhor escolha, ao passo que se houver uma preferência de atuação no longo prazo, a segunda apresenta um melhor encaixe. Em qualquer um dos casos, o que realmente importa é que movimentos transitórios oriundos de choques de oferta não entram no conceito de núcleo, além de que essa medida terá uma volatilidade menor do que o índice cheio justamente por esse motivo.

Por outro lado, a segunda linha de pensamento buscará definir o núcleo pelo caminho de uma inflação generalizada. Aqui, a inflação é vista como uma junção entre um termo generalizado (núcleo), e um outro termo de ruído, que representa, em sua maioria, choques de oferta. É importante salientar que esse ruído denota mudanças nos preços relativos que podem ser de outra natureza, e que, portanto, poderia ter um impacto diferenciado no índice de inflação. Porém, ao se assumir que o que prepondera são choques de oferta, as duas definições de núcleo apresentadas representam basicamente a mesma coisa, pois existirá um componente duradouro (generalizado x persistente) e um outro transitório. Ainda assim, é mais comum que os Bancos Centrais utilizem esta tese de inflação generalizada para tratar do núcleo; em outras palavras “In practice, most central bankers tend to talk about core inflation in such terms, defining core inflation as the aggregate inflation excluding a variety of items whose price movements are deemed likely to distort or obscure the more general trend of other prices.” (Roger, 1998).

Com a definição de núcleo apresentada, o que verdadeiramente parece importar é a forma de se mensurar essa variável. É pertinente dizer que não existe um consenso sobre qual é a melhor medida de núcleo, e que por esse motivo é natural que diferentes países possam usar métricas diferentes entre si<sup>1</sup>.

---

<sup>1</sup> Um apanhado de diferentes medidas de núcleo de inflação utilizadas por Bancos Centrais ao redor do mundo pode ser observado em <https://www.bcb.gov.br/htms/relinf/port/2000/06/ri200006b4p.pdf>

Como revelam Castelar e Santos (2016), é possível agrupar as medidas de núcleo em três grupos: (i) de séries temporais, que refletem a inflação persistente; (ii) de corte transversal, que reflete a inflação generalizada; e (iii) de medidas de painel, que é uma junção das duas, mas que não será explorado aqui. Sobre o primeiro grupo, podem ser usadas séries temporais univariadas ou multivariadas, ambas com o objetivo de captar a persistência na inflação. No modelo univariado geralmente se usa alguma técnica de suavização para conseguir filtrar a série e encontrar o componente de persistência através da frequência. Contudo, Roger (1998) ressalta que esse método não é robusto pelo fato de que, por exemplo, em um caso de ajuste sazonal, o modelo só consegue assimilar choques transitórios que tenham uma determinada regularidade e magnitude; se for um choque atípico, o modelo deixa passar.

No caso multivariado, o que acontece é que a forma de se identificar o núcleo é decompondo a inflação entre um termo que afeta o produto real de forma perene, e aquele que não tem efeito no médio e longo prazo (núcleo). Contudo, um problema dessa abordagem é permitir que todo choque transitório seja de um só tipo, além de outras questões técnicas.

Em contrapartida, as medidas de corte transversal são mais simples e comuns de serem aplicadas. Como já mencionado, a mais famosa dessas métricas é o núcleo por exclusão. Nesse caso, o que acontece é que determinados itens do índice de preço são calculados novamente (*re-weighting*) com um novo peso, ou até mesmo excluído, se eles forem avaliados como sendo oriundos de choques de oferta. Consequentemente, os itens que são entendidos como sendo menos perceptíveis a esses choques acabam por receber um peso maior nessa nova avaliação. No caso brasileiro, entre as medidas de núcleo mais comumente acompanhadas, o IPCA-EX0 e o IPCA-EX1 trabalham com essa ideia de subtrair alguns elementos do IPCA cheio. Assim, o primeiro deles exclui o item Alimentação no Domicílio e Monitorados, enquanto o segundo exclui combustíveis e alguns dos subitens de Alimentação no Domicílio<sup>2</sup>. Por outro lado, o IPCA-DP só reduz o peso dos itens com maior volatilidade no índice, sem precisar excluí-los.

Dentre os potenciais problemas que esse tipo de abordagem apresenta, talvez o pior seja a possibilidade de que o núcleo tenha uma tendência diferente do índice agregado. Esse impasse acontece se a série que foi manipulada na hora de re-ponderar tiver originalmente uma tendência diferente do IPCA, pois nesse caso o que se verifica na prática é que, na hora da manipulação, foram excluídos itens que faziam parte da tendência da série, ou seja, que não eram apenas componentes transitórios (Roger, 1998). No limite, o que pode ocorrer se a subtração de itens

---

<sup>2</sup> Uma breve explicação sobre os 5 núcleos que são acompanhados com maior frequência pelo Banco Central do Brasil pode ser vista em: [https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/Novas\\_medidas\\_nucleo\\_inflacao.pdf](https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/Novas_medidas_nucleo_inflacao.pdf)

voláteis não for bem feita, é que a medida de inflação subjacente terá menos utilidade para acompanhar a tendência do que a própria inflação cheia; nesse caso, o uso dessa medida de núcleo específica seria inútil.

Via de regra, justamente por todas as medidas de núcleo apresentarem defeitos e não terem uma consistência metodológica precisa, o mais habitual é que os Bancos Centrais acompanhem um conjunto desses indicadores (da Silva e Figueiredo, 2014). Sendo assim, no Brasil, ao todo são cinco as principais medidas de acompanhamento do núcleo da inflação, sendo que é absolutamente normal o BCB divulgar novas medidas de tempos em tempos se a instituição achar que é prudente do ponto de vista prático na análise da inflação. Foi com esse tipo de raciocínio que, em 2018, foram introduzidos dois novos índices, o IPCA-EX2 e o IPCA-EX3. Mesmo divergindo da média dos outros núcleos durante grande parte do tempo, e tendo um desempenho aquém em alguns testes econométricos realizados, eles mostraram-se úteis em uma análise que leva em consideração o ciclo econômico (BCB, 2018).

Em relação à capacidade de conseguir prever a tendência, da Silva & Figueiredo (2014) fazem um estudo detalhado para afirmar que, pelo menos para o período analisado, até mesmo novas medidas que são apresentadas ou modificadas não apresentam uma diferença que possa ser realmente considerável, mesmo que essas novas medidas eventualmente acabem por substituir uma outra que já era usada há mais tempo, e apresenta um resultado piorado relativamente. Sendo mais rigorosos, Castelar & Santos (2016) mostram que as medidas usadas atualmente pelo BC não são úteis para fazer previsão do núcleo fora da amostra, e propõem uma nova medida. Considerando três critérios (ausência de viés, dinâmica de ajustamento e capacidade preditiva), o que eles observaram para o período de análise é que, embora alguns dos núcleos regularmente usados passassem em um critério ou outro, nenhum deles foi capaz de passar nos três testes.

Nesse sentido, se o que se busca é uma medida de núcleo que tenha uma capacidade preditiva muito forte para o caso brasileiro, talvez seja interessante que se adote essa nova métrica. No entanto, ao que parece, o objetivo do Banco Central não é necessariamente prever a inflação para determinado horizonte utilizando algum tipo de núcleo, mas simplesmente observar a tendência dos preços ao longo do tempo e conseguir perceber choques temporários no índice agregado, além de auxiliar na condução da política monetária, algo que aparentemente tem sido feito. Como já existe um conhecimento prévio geral por parte das instituições monetárias de que cada medida apresenta certos problemas, e que não existe uma que seja melhor sempre para todos os critérios ao mesmo tempo em todos os horizontes temporais, o

uso de um conjunto de métricas é sempre a melhor escolha, já que o desempenho de cada uma delas pode variar dependendo da natureza do choque.

Ainda assim, antes da crise de 2008, alguns economistas monetários (como Woodford, 2003, e Goodfriend, 2007, *apud* Ruch et al. 2019, p. 2) haviam chegado em uma certa convergência de que o núcleo da inflação seria uma âncora nominal melhor do que a inflação efetiva para a condução da política monetária, e que, portanto, deveria haver uma meta para ele. Dessa forma, há alguns autores que partem desse pressuposto para se aventurarem na tarefa de tentar prever o núcleo, ao invés do índice cheio. Então, Ruch et al. (2019) optam por realizar esse trabalho para o caso sul africano e conseguem obter bons resultados no que diz respeito ao uso de modelos VAR se comparados a um passeio aleatório e outros modelos de *benchmark* para todos os períodos testados.

Ou seja, a despeito de o núcleo ser uma variável de bastante importância para o acompanhamento da política monetária, não se observam muitos trabalhos no sentido de prevê-lo.

#### **1.4 A importância da previsão**

Existem alguns motivos que enfatizam a necessidade de que haja a projeção de variáveis macroeconômicas, neste caso, a inflação. Podemos dividir essa importância em três situações: na coordenação da política monetária; na tomada de decisão das famílias e no estabelecimento de contratos para o setor privado. Começaremos frisando a relevância na política monetária.

De acordo com Svensson (2005), pelo fato de as ações de política monetária funcionarem com uma certa defasagem, é preciso que elas sejam *forward-looking* no sentido de conseguirem alterar o estado da economia no futuro, e para isso é fundamental que essas ações sejam baseadas em previsões. Assim, sabe-se que os funcionários de bancos centrais trabalham fazendo projeções de uma série de variáveis, tanto exógenas, quanto endógenas, enquanto os *policymakers* recebem análises futuras de variáveis mais importantes, como o PIB e a inflação. Na política monetária moderna, as chamadas '*targeting rules*' funcionam como um pré-requisito para a sua boa condução, e mesmo que sejam simples, podem fazer com que a política monetária seja robusta e funcione bem em uma gama de situações.

Nessa perspectiva, a ideia de *inflation targeting* (regime de metas de inflação) foi adotada por diversos Bancos Centrais no mundo e, além de auxiliar na condução da política monetária, também aumenta a transparência, uma vez que existirá uma maior comunicação dos

polycymakers com o público e o mercado. Isso é feito por meio de divulgações dos relatórios de inflação, nos quais as projeções são feitas considerando toda a conjuntura interna e externa. No caso brasileiro, o BCB diz explicitamente que para atingir seu principal objetivo de estabilidade de preços, é fundamental que isso seja feito pelo regime de metas<sup>3</sup>. Ou então, como a própria instituição afirma, “O cumprimento do objetivo de assegurar a estabilidade de preços se dá por meio do regime de metas para a inflação, as quais são estabelecidas pelo Conselho Monetário Nacional.” (BCB, 2024). Como explica Svensson (1997), uma das vantagens da adoção do regime de metas é a busca por uma inflação que seja baixa e estável, pois isso pode ajudar a diminuir (ou eliminar) o viés inflacionário, além da ancoragem das expectativas. Como se sabe, uma inflação mais alta é nociva tanto para o ambiente institucional, minando a confiança dos agentes, assim como por ter efeitos regressivos, impactando os mais necessitados; ou seja, a adoção de uma meta que seja baixa ajuda a evitar essas distorções. Sem entrar em detalhes, também se sabe que o regime de metas apresenta certos problemas por conta do *delay* da política monetária, no sentido de que a inflação de um período é fruto de fatores pretéritos e de elementos que estão fora do controle do BC. Assim, pode ser difícil a implementação desse tipo de regime.

Voltando à questão da previsão da inflação, uma das conclusões que Svensson (1997) apresenta é que em um cenário de inflation targeting, a própria previsão de inflação divulgada pelo Banco Central acaba se tornando a sua meta intermediária. Por exemplo, se a política monetária só consegue afetar a inflação com um atraso de cerca de 2 anos, a solução ótima que o autor chega é que a taxa de juros (*repo rate*) de hoje deve ser tal que consiga igualar a taxa de inflação prevista de dois anos com a meta. Ou seja, a taxa de juros deve igualar

$$\Pi_{t+2|t} = \Pi^* \quad (2)$$

onde  $\Pi_{t+2|t}$  é a previsão para a taxa de inflação no ano ‘t+2’, dada a informação disponível em ‘t’. E  $\Pi^*$  é a meta de inflação. Assim, havendo algum desvio entre essas duas taxas, o ajuste será feito por mudanças no instrumento de política monetária (taxa de juros). Já que o BC não consegue afetar a inflação contemporânea, o máximo (e melhor) que ele consegue fazer é garantir que a previsão de inflação seja igual à meta do regime. Uma meta intermediária é mais fácil de ser controlada pelo Banco Central, além de ser mais fácil de ser observada na

---

<sup>3</sup> Os últimos relatórios de inflação do Banco Central do Brasil podem ser vistos em: <https://www.bcb.gov.br/publicacoes/ri>

prática e facilitar na comunicação com os agentes. É nessa direção que a divulgação dos relatórios de inflação ajuda a melhorar o monitoramento do público com relação às previsões que são feitas pelo BC. Isso é relevante para que se veja possíveis desvios em relação à meta, e como os especialistas do BC estão enxergando a evolução do cenário nacional e internacional.

Mais especificamente, Svensson enumera alguns motivos que fazem com que a previsão de inflação seja a meta que o BC venha a seguir. Pode-se destacar que é mais fácil controlar a previsão para a taxa de inflação futura do que a taxa de inflação futura *per se*, pois o efeito do instrumento na previsão de inflação é o mesmo que na inflação média, e a variância da previsão é menor. Além disso, é mais fácil observar a projeção da inflação, porque a previsão para  $t+2$  dado  $t$  é uma variável que é observável contemporaneamente em  $t$ , já que só depende da informação do período atual. Ou seja, não é preciso que se chegue em  $t+2$  para visualizá-la. Como ressaltado, o público interessado consegue ter acesso pelos relatórios. Outras razões a se colocar dizem respeito ao surgimento de incentivos aos funcionários do Banco Central na tarefa de entenderem melhor o funcionamento dos modelos que melhor preveem a inflação. Por fim, a tarefa de prever as ações de política monetária serão semelhantes à previsão de inflação, na medida em que as informações necessárias para se prever a PM serão as mesmas usadas para prever a inflação.

No caso das famílias, a grande importância surge com o papel das expectativas, especialmente as de curto prazo. Segundo Adrian (2023), durante muito tempo houve uma percepção de que, pelo fato de as expectativas de inflação das famílias serem voláteis e dispersas, elas não teriam tanta utilidade para a política monetária. Porém, na verdade, o que se observou posteriormente é que essa dispersão era útil para o Banco Central em termos de informação. A distribuição das expectativas contém ingredientes que ajudam o BC na previsão da inflação. Por exemplo, mudanças na distribuição das expectativas de curto prazo (1 ano) das famílias podem ser usadas como uma boa variável para prever a inflação de 1 ano.

Ainda de acordo com Adrian (2023), e como também pode ser visto em Andrade et al. (2020), as expectativas das famílias sobre a inflação acabam por influenciar a demanda por conta de um aumento no consumo. Ou seja, o que os autores descobriram é que as famílias antecipam a compra de bens duráveis se elas esperam que haverá um aumento de preço, ao passo que as famílias que acreditam em uma estabilidade no nível de preços mantêm a mesma propensão marginal a consumir (vale salientar que, no caso da pesquisa feita por Andrade et al., eles só têm informações sobre bens duráveis). Ademais, as decisões de comprar ou alugar um imóvel são determinadas em parte pela expectativa do que as famílias esperam para os preços futuros. Basicamente, se acham que os preços subirão no futuro, preferem comprar um

imóvel a alugar; da mesma forma, a negociação sobre a taxa de juros (fixa ou flutuante) incidente em um contrato de hipoteca dependerá das expectativas de inflação. Por último, essas expectativas também acarretam em tentativas de negociação de salários, horas trabalhadas ou mudança de emprego. Apesar de algumas dessas conclusões parecerem óbvias, elas têm implicações na forma com que a política monetária deverá atuar.

Por conta das expectativas de curto prazo das famílias, a autoridade monetária precisa estar atenta para que não haja o risco de elas sofrerem uma desancoragem. Atualmente, muito pelo fato de as expectativas de longo prazo estarem bem ancoradas em muitos lugares, é necessário que os BC's estejam bem vigilantes sobre as de curto. Em um cenário com altas taxas de inflação, existe a chance de ocorrer essa desancoragem, e conseqüentemente uma dificuldade para baixar as expectativas, já que elas respondem muito mais a aumentos do que cortes, sem contar que as famílias se informam mais por outras fontes sobre economia do que propriamente o Banco Central (Adrian, 2023). Assim, a solução é uma postura rígida do BC no trato da política monetária, para que consiga trazer tanto a inflação quanto as expectativas para baixo, e conseguir manter uma boa credibilidade. No caso brasileiro, por possuir um bom arcabouço de PM já estabelecido, esse trabalho é facilitado.

No que diz respeito à importância da previsão da inflação para as firmas, é importante que os empresários tenham algum grau de previsibilidade sobre a evolução dos preços já que isso afetará as decisões sobre investimentos, contratação de novos funcionários e negociação de salários. Uma das formas usadas para se entender como as empresas reagem nos momentos de mudança de expectativa é por meio de pesquisas qualitativas. Nesse sentido, Coibion, Gorodnichenko e Ropele (2018) fazem esse exercício para empresas italianas; já Kaihatsu & Shiraki (2016) o fazem para empresas japonesas. O primeiro grupo de autores observou que, quando as firmas tinham um aumento nas suas expectativas de inflação em 1%, isso implicava um aumento transitório (3 meses) nos preços de seus produtos, além de uma queda na demanda por emprego que durava mais de 1 ano, e um aumento na busca por crédito que durava cerca de 1 ano. Entretanto, essa dinâmica pode variar um pouco para o caso de países que estão no limite inferior da taxa de juros (*effective lower bound*). No caso dos autores japoneses, foi feita uma análise entre as expectativas de inflação e o pagamento de salários aos funcionários de uma empresa por meio de uma causalidade de Granger<sup>4</sup>. O que se concluiu foi que um aumento

---

<sup>4</sup> Segundo Bueno (2011), diz-se que, se o escalar  $y$  ajuda a prever o escalar  $z$ , então  $y$  Granger-causa  $z$ . O autor cita o exemplo das taxas de juros, no qual o teste de causalidade de Granger é usado para verificar se a meta de juros definida pelo BC Granger-causa a taxa de mercado, enquanto o oposto não é válido. Esse exercício também foi feito por John Taylor para o caso americano.

das expectativas de inflação de curto prazo das firmas leva à uma queda dos salários e dos lucros, ao passo que um aumento das expectativas de médio e longo levam a aumento dos salários. A interpretação que eles fizeram foi de que é um pouco mais difícil repassar o aumento de alguns custos para os preços dos produtos no curto prazo, o que ocasiona uma diminuição nas margens de lucros e, conseqüentemente, salários menores. No entanto, em períodos mais longos, para que haja um crescimento equilibrado entre preços e salários, o último deve aumentar ao longo do tempo; também é ressaltado a maior importância relativa de horizontes maiores pois pode acontecer de as expectativas de curto prazo das firmas não impactarem as a dinâmica salarial.

### **1.5 Alguns potenciais problemas ao fazer previsão**

De acordo com Stock & Watson (1999), ao analisarem a série da inflação americana, é possível que sejam feitos dois tipos de conclusões, dependendo do ponto de vista do analista, quais sejam: que a inflação americana ficou, ao mesmo tempo, mais fácil e mais difícil de se prever. Tornou-se mais fácil de se prever pelo fato de ter ocorrido uma queda na volatilidade da série em comparação a períodos anteriores (onde tiveram episódios de choques de diferentes naturezas), além de uma queda no RMSE dos modelos mais básicos. Porém, ao mesmo tempo, as técnicas que eram mais usadas naquela época para previsão pouco diferiam de um simples modelo ingênuo, ou seja, não necessariamente conseguiam resultados melhores, o que tornava mais difícil a tarefa de previsão.

Embora a análise seja feita para os Estados Unidos, talvez um resultado generalizável no artigo de Stock & Watson seja o fato de que quase necessariamente o desempenho dos modelos empregados em um estudo dependerá da característica da série. Como os autores a dividem em dois momentos, tornava-se mais claro perceber que alguns métodos performavam melhor que o benchmark (AR) em um certo recorte, mas depois apresentavam um resultado grosseiramente pior, como era o caso da Curva de Phillips. Notadamente, no período conhecido como Grande Moderação (*Great Moderation*) nos países desenvolvidos, houve um arrefecimento das flutuações em um conjunto de variáveis macroeconômicas, o que de certa forma facilita na análise e previsão dessas séries. De certa forma, para o caso brasileiro, fazer o exercício de previsão em um cenário pós Plano Real também tende a ajudar no sentido de conseguir resultados mais satisfatórios.

No entanto, como ressalta Tulip (2009), a redução da volatilidade de uma série não implica que ela será mais previsível (ou menos incerta), que é o que importa verdadeiramente para a maioria dos agentes, pois quando há incerteza, os indivíduos costumam tomar decisões erradas ou equivocadas sobre determinado assunto, como investimentos mal feitos ou *hedge* para se protegerem de uma situação que jamais chegou perto de acontecer. No estudo de Tulip, o que se constatou foi que de fato houve uma redução da incerteza da inflação e PIB americanos, embora significativamente menores do que a redução da volatilidade. A razão para isso é que o componente previsível (*predictable*) das flutuações foi se reduzindo ao longo dos anos. Por definição, o autor apresenta esse componente previsível como sendo a diferença entre a variância e o EQM; e como a variância diminuiu no período de Grande Moderação, isso acaba por explicar um pouco o motivo dessa queda na previsibilidade.

Ainda sobre os EUA, uma outra dificuldade ratificada por El Shagi, Giesen e Jung (2014) diz respeito ao conjunto de informações disponíveis para aqueles que tentam fazer previsão. Como foi evidenciado por eles, as previsões de PIB e inflação feitas pelo FED (*Greenbook forecasts*) normalmente superavam àquelas de instituições privadas (tinham menor RMSE), especialmente nos períodos em que a incerteza era maior. Segundo os autores, essa melhor performance relativa surge por conta de uma vantagem de informação que o Banco Central tem sobre outros players, e que pode ser explicada em parte por um maior detalhamento que o BC americano tem sobre o futuro da taxa dos Fed Funds, embora existam outras explicações diferentes dessa.

É fato que as projeções do *Greenbook* só ficam disponíveis após um período de 5 anos (ainda que seja possível ter acesso às projeções do FOMC a cada três meses), o que em tese dificultaria a tarefa de previsão de instituições privadas (e outros entusiastas) já que não teriam acesso em tempo real ao que o Fed está vendo. No entanto, por conta de uma busca no aumento da transparência por parte dos BC's, esse gap informacional e, conseqüentemente, de performance, tem caído ao longo do tempo. Além disso, com a maior quantidade de dinheiro investida em previsão por parte de bancos e empresas, o desempenho dessas corporações em relação ao *Greenbook* também melhorou. Outro fator a se destacar é que a partir do momento em que a volatilidade do ciclo econômico diminuiu (década de 1980), os resultados de previsão não ficaram mais tão discrepantes quanto eram anteriormente. Mesmo assim, com essas melhorias, ainda é extremamente difícil conseguir resultados similares ou superiores ao Fed na previsão do PIB e inflação (El Shagi, Giesen e Jung, 2014).

Para o caso brasileiro, no que diz respeito ao IPCA, algumas considerações são importantes de se fazer. No artigo de Medeiros, Vasconcelos e Freitas (2016), assim como em

outros sobre o tema, fica constatado que a inflação do Brasil tem um alto grau de autocorrelação, implicando, na prática, que modelos do tipo auto regressivo são bons preditores. De fato, o que eles observaram para o período analisado é que o IPCA pode ser decomposto em um componente inercial, e variações na dívida pública. Eles chegaram a essa conclusão ao usarem modelos de encolhimento que escolhem as melhores variáveis explanatórias e, na maioria das vezes, elas estavam relacionadas com a dívida e moeda (como M3 e M4). Nesse sentido, para o curto prazo, esse grupo de variáveis ajuda na previsão, e o AR tem sua parcela de ajuda por conta do componente inercial. Ainda se tratando da seleção feita por modelos de encolhimento, é interessante perceber que a Curva de Phillips tende a não ser uma boa ferramenta para a previsão porque variáveis como desemprego e produção não se mostraram de grande importância na seleção.

Porém, no longo prazo, um resultado que pode impactar negativamente no exercício de previsão é saber que a maioria das variáveis de natureza econômica preveem muito mal o IPCA, ou seja, de que os próprios índices de inflação têm um desempenho melhor. Em outras palavras, a inflação medida mês a mês para o caso brasileiro só pode ser aproximada pelo próprio componente inercial, captado pelo AR, enquanto variáveis macroeconômicas se tornam inúteis. Nessa perspectiva, duas questões podem ser levantadas: (i) a primeira delas é compreender se essa dinâmica ainda é válida para os anos mais recentes, o que provavelmente deve se confirmar; (ii) a segunda é tentar entender se os novos modelos de *machine learning* conseguirão de fato ter alguma influência sobre esse resultado, isto é, se eles terão alguma capacidade de ter um desempenho superior a um modelo ingênuo para horizontes maiores, principalmente por se saber que eles melhoram a performance de acordo com o avançar dos períodos.

## **CAPÍTULO II – METODOLOGIA**

O objetivo deste capítulo é apresentar os dados e os principais aspectos da metodologia utilizada neste trabalho para a previsão do IPCA. Para tal, o capítulo será dividido três partes: a base de dados utilizada, com os devidos detalhes necessários; o conjunto de técnicas que foram usadas para previsão no trabalho, e o critério de comparação entre os modelos. Não haverá uma extensa explicação sobre variáveis e métodos, sendo apresentado apenas o suficiente para a compreensão do leitor.

### **2.1 Base de dados**

Como o principal objetivo desta dissertação é fazer previsão da inflação brasileira, vale destacar de início a variável target escolhida para essa tarefa. Dentre os diversos índices de inflação existentes, o IPCA cheio foi o escolhido por conta de sua importância relativa em relação aos demais. Este é o índice oficial usado pelo governo para acompanhamento do regime de metas de inflação, é calculado pelo IBGE, e abarca as famílias que têm rendimentos mensais de 1 a 40 salários mínimos nas principais metrópoles do país. Um ponto interessante ressaltado por Medeiros, Vasconcelos e Freitas (2016) é que o IPCA é uma variável com um alto grau de autocorrelação, o que significa dizer que modelos do tipo autorregressivo tendem a ter sucesso na previsão desse índice. Esse fato também foi percebido por outros autores que tentaram prever a inflação brasileira anteriormente. Nesse sentido, a inclusão de um modelo AR provavelmente será de grande valia.

No que diz respeito às variáveis escolhidas para serem as preditoras, a melhor escolha foi novamente seguir o que já é bem documentado na literatura em trabalhos de previsão de inflação, isto é, escolher variáveis dos seguintes grupos: índices de preços, agregados monetários, setor financeiro, setor externo, setor público e produção. Na tabela 1, apresentamos as variáveis utilizadas, suas fontes e frequência. Além disso, indicamos qual o tipo de transformação (se alguma) foi realizada para que a variável se torne estacionária. Vale ressaltar que, para utilização dos métodos propostos, especialmente os de ML, as séries devem ser  $I(0)$ , isto é, estacionárias.

Todas as variáveis passaram pelo teste de Augmented Dickey-Fuller (ADF) para verificar se eram estacionárias ou não. Após isso, as seguintes transformações foram aplicadas nas variáveis: (0) sem transformação; (1)  $\Delta x_t$ ; (2)  $\Delta^2 x_t$ ; (3)  $\Delta \ln x_t$ ; (4)  $\Delta^2 \ln x_t$ . Essas transformações foram feitas para ter resultados comparáveis ao trabalho de Araújo & Gaglianone (2023).

**Tabela 1 – Variáveis usadas e suas informações**

| <b>Série</b>                      | <b>Frequência</b> | <b>Fonte</b> | <b>Transformação</b> |
|-----------------------------------|-------------------|--------------|----------------------|
| IPCA                              | mensal            | IBGE         | 0                    |
| IPCA alimentos                    | mensal            | IBGE         | 0                    |
| IPCA habitação                    | mensal            | IBGE         | 0                    |
| IPCA transportes                  | mensal            | IBGE         | 0                    |
| IPCA saúde                        | mensal            | IBGE         | 0                    |
| IPCA educação                     | mensal            | IBGE         | 0                    |
| IPCA serviços                     | mensal            | IBGE         | 0                    |
| IPCA 15                           | mensal            | IBGE         | 0                    |
| IPCA núcleo EX0                   | mensal            | BCB          | 0                    |
| IPCA núcleo EX1                   | mensal            | BCB          | 0                    |
| IPC                               | mensal            | FIPE         | 0                    |
| INPC                              | mensal            | IBGE         | 0                    |
| IGP-M                             | mensal            | FGV          | 0                    |
| IGP-DI                            | mensal            | FGV          | 0                    |
| INCC-DI                           | mensal            | FGV          | 0                    |
| IGP 10                            | mensal            | FGV          | 0                    |
| Base monetária                    | mensal            | BCB          | 3                    |
| M1                                | mensal            | BCB          | 3                    |
| M2                                | mensal            | BCB          | 4                    |
| M3                                | mensal            | BCB          | 4                    |
| M4                                | mensal            | BCB          | 4                    |
| Compulsório                       | mensal            | BCB          | 1                    |
| SELIC                             | anual             | BCB          | 1                    |
| TJLP                              | anual             | BCB          | 1                    |
| Reservas internacionais           | mensal            | BCB          | 4                    |
| IBC-BR                            | mensal            | BCB          | 3                    |
| PIB                               | trimestral        | IBGE         | 4                    |
| Total de vendas varejo            | mensal            | IBGE         | 0                    |
| Taxa de câmbio nominal            | diária            | BCB          | 3                    |
| EMBI+                             | diária            | IPEADATA     | 3                    |
| CRB                               | mensal            | STOOQ        | 3                    |
| Petróleo Brent                    | diária            | IPEADATA     | 3                    |
| Total de exportações              | mensal            | Comex Stat   | 3                    |
| Total de importações              | mensal            | Comex Stat   | 3                    |
| Índice de confiança do consumidor | mensal            | Fecomercio   | 1                    |
| IBOVESPA                          | diária            | B3           | 3                    |
| Produção de veículos              | mensal            | Anfavea      | 3                    |
| DLSP                              | mensal            | BCB          | 2                    |
| DLSP / PIB                        | mensal            | BCB          | 2                    |
| Resultado primário                | mensal            | BCB          | 1                    |
| Resultado primário / PIB          | mensal            | BCB          | 1                    |

A base de dados acima engloba um período de janeiro de 2006 até dezembro de 2017. A aplicação das metodologias de machine learning exige que parte dos dados sejam usados para “treinamento” do modelo. Os dados de treino representam 70% das observações, enquanto os de teste são os outros 30%. Essa divisão costuma ser de praxe em trabalhos de machine learning, mesmo que não seja uma regra, e possa variar. Ela costuma ser o suficiente para o modelo aprender bem e ter capacidade de generalizar para novos dados. Desse modo, a base de dados até 2017 será repartida nessa proporção de forma aleatória para conseguir um melhor *fit* dos modelos, sendo que a previsão fora da amostra será realizada para os anos de 2018 e 2019. Para isso, foram usados intervalos de previsão para 6, 9, 12 e 24 meses ( $h = 6, \dots, 24$ ).

Os códigos computacionais para este trabalho foram feitos na linguagem de programação R, por meio da IDE Rstudio. Para o modelo autorregressivo e o autorregressivo de médias móveis, fez-se uso do pacote *forecast* e da função Arima. Já para os modelos de machine learning, algumas vezes foram usados mais de um pacote para testar diferentes especificações/tuning dos parâmetros. Assim, para o KNN foram usados os pacotes *caret* e *tidymodels*, para o SVM também se usou o *caret*, além do pacote *e1071*. Por fim, para o LightGBM, novamente foi usado o pacote *tidymodels*.

É interessante salientar que uma das inspirações para esta dissertação foi o trabalho de Araújo & Gaglianone (2023), em que os autores também fazem previsão do IPCA. Os autores usam uma gama de modelos de machine learning (geralmente tidos como os mais conhecidos), além de outros tipos de modelos que não serão abordados aqui. O nosso trabalho se diferencia de Araújo e Gaglianone por termos optado por outros modelos de machine learning, notadamente KNN, SVM e LightGBM. O primeiro deles foi escolhido por não ser incomum de apresentar resultados razoáveis em dados não-lineares e ser de fácil implementação. O SVM teve um bom desempenho ao prever a inflação da Inglaterra no artigo de Chakraborty & Joseph (2017), e por isso foi escolhido aqui, na tentativa de verificar se também seria um bom predictor para o caso brasileiro. Por fim, também foi optado pelo Light GBM por conta de suas boas previsões em competições de ciência de dados (como no Kaggle) e suas vantagens computacionais. Dessa forma, podemos verificar a eficiência desses novos modelos em relação ao benchmark, e adicionarmos novidades em relação ao trabalho daqueles autores.

## 2.2 Modelos

Para se atingir o objetivo desta dissertação, foram usados 5 modelos, divididos da seguinte forma: 2 modelos de séries temporais, um modelo autorregressivo (AR, será o benchmark) e um modelo autorregressivo de média móvel (ARMA); e 3 modelos de machine learning, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) e LightGBM<sup>5</sup>.

### 2.2.1 Modelos de séries de tempo

#### 2.2.1.1 Modelo autorregressivo (AR) – *benchmark*

É comum que em um trabalho de previsão seja utilizado algum modelo “ingênuo” (naive) que faça estimações mais simples (mas não necessariamente inferiores). Geralmente, usa-se um modelo do tipo autorregressivo, que explica a variável por suas próprias defasagens. Nesta dissertação, esse modelo ingênuo terá a especificação de um modelo autorregressivo.

De acordo com Arruda, Ferreira e Castelar (2011), um modelo autorregressivo tem uma boa performance se a inflação for uma função linear de suas observações passadas, representando um componente de memória inflacionária. Segundo Bueno (2011), um modelo autorregressivo de ordem ‘p’ pode ser escrito como:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (3)$$

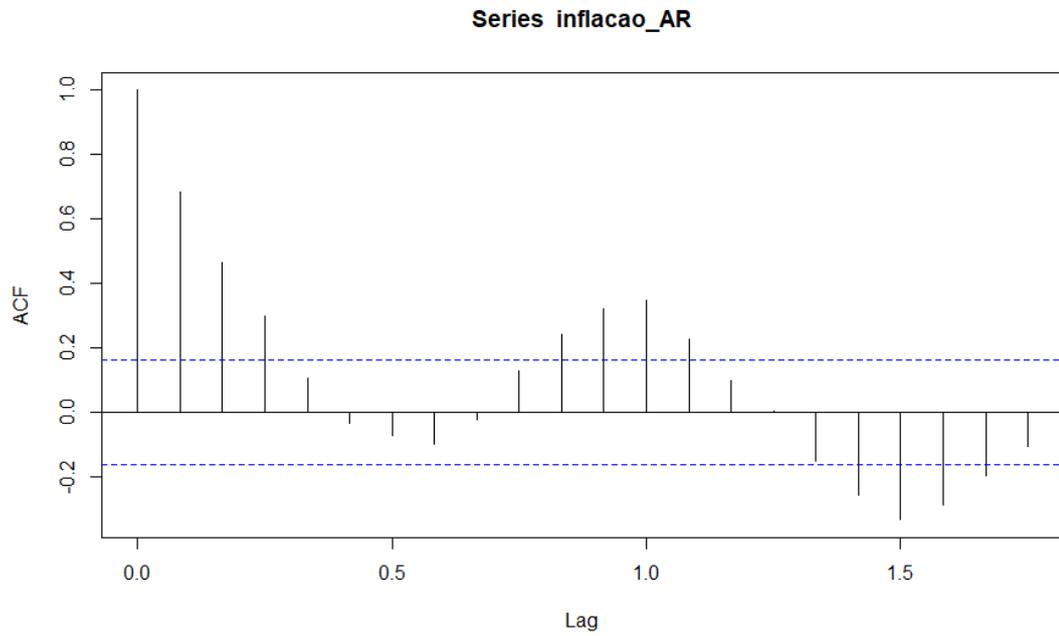
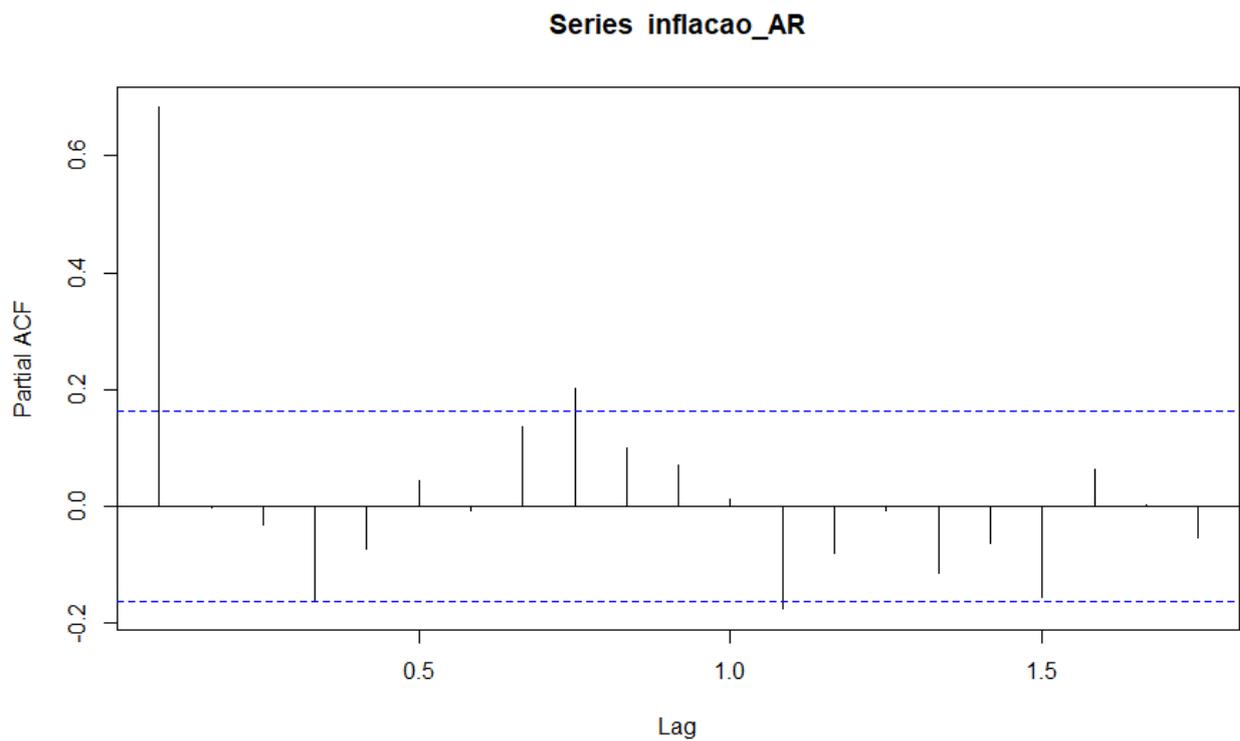
Sendo o termo de erro  $\varepsilon_t \sim N(0, \sigma^2)$ . O método para definir a ordem do processo é usar os critérios de informação (AIC/BIC/HQ) e ver qual deles possui uma estatística menor, ou então pela análise da função de autocorrelação (FAC) e função de autocorrelação parcial (FACP). A previsão de ‘h’ períodos a frente é dado por:

$$\widehat{y}_{t+h} = \hat{c} + \hat{\varphi}_1 \hat{y}_{t+h-1} + \hat{\varphi}_2 \hat{y}_{t+h-2} + \dots + \hat{\varphi}_p \hat{y}_{t+h-p} \quad (4)$$

Os gráficos 1 e 2 abaixo mostram a FAC e FACP, que ajudaram a descobrir a ordem do modelo, aplicado ao IPCA brasileiro do período analisado.

---

<sup>5</sup> Na primeira versão deste trabalho, também foi estimado um MIDAS para prever o IPCA. Porém, ele foi retirado por questões metodológicas.

**Figura 1 – Função de autocorrelação – IPCA****Figura 2 – Função de autocorrelação parcial – IPCA**

Observando ambos os gráficos, o primeiro ponto a se destacar é a confirmação do fato apontado por Medeiros, Vasconcelos e Freitas (2016) sobre o IPCA ter um grande nível de autocorrelação, chegando a ultrapassar 0.6 no primeiro *lag* e sendo estatisticamente significativa até o terceiro *lag* (gráfico 1). O gráfico 2 sugere um modelo autorregressivo de primeira ordem, e que certamente foi testado juntamente com outros modelos até o de terceira ordem. Ao usar a função *auto.arima* no software Rstudio, a recomendação é a de um modelo de primeira ordem, com um AIC de -14.29, e um BIC de -4.92. Portanto, a princípio, esse será o modelo escolhido como benchmark, mas, como veremos mais adiante, serão adicionados os modelos AR (2) e AR (3). Também foi incluído o próprio IPCA com 3 defasagens como variável preditora.

### 2.2.1.2 Processo Autorregressivo de Médias Móveis - ARMA (4,3)

Como o nome sugere, um processo ARMA é a mistura de um modelo autorregressivo (AR) com um modelo de médias móveis (MA). Novamente, usando a notação de Bueno (2011), podemos escrever um ARMA (4,3) com a seguinte representação:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \varphi_4 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} \quad (8)$$

Como  $\varepsilon_t \sim N(0, \sigma^2)$ , podemos escrever a previsão de um modelo ARMA (4,3) em ‘h’ períodos a frente da seguinte maneira:

$$\hat{y}_{t+h} = \hat{c} + \hat{\varphi}_1 \hat{y}_{t+h-1} + \hat{\varphi}_2 \hat{y}_{t+h-2} + \hat{\varphi}_3 \hat{y}_{t+h-3} + \hat{\varphi}_4 \hat{y}_{t+h-4} \quad (9)$$

Vale salientar que a inclusão desse modelo especificamente se deve ao fato de o Banco Central do Brasil usar essa mesma especificação para fazer projeções sobre o IPCA em um intervalo entre 2001 e 2018, isto é, um espaço temporal que cobre grande parte do período aqui analisado<sup>6</sup>.

---

<sup>6</sup> O estudo do BCB que usa o referido modelo pode ser acessado em: [https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/Modelo\\_de\\_vetor\\_autorregressivo\\_com\\_ancoragem\\_de\\_longo\\_prazo.pdf](https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/Modelo_de_vetor_autorregressivo_com_ancoragem_de_longo_prazo.pdf)

## 2.2.2 Modelos de machine learning

Nos últimos anos, com os avanços de técnicas computacionais, diversos novos métodos têm sido incorporados para a análise de dados que vão desde simples tarefas (como a criação de um filtro de *spam*) até grandes conteúdos econômicos e financeiros. Nesse sentido, o uso das técnicas de *machine learning* tem ajudado cientistas de diversas áreas a apresentarem resultados satisfatórios em muitas pesquisas. Contudo, não necessariamente essas técnicas têm crescido entre econométricos e estatísticos por conta de seus objetivos de análise. Como destacam Einav & Levin (2014), esses profissionais estão muito mais interessados em algum tipo de relação causal entre os preditores e a variável dependente, ou alguma tentativa de avaliar determinado tipo de política. Ou colocado de outra forma, o que ocorre é que muitas vezes um economista está interessado em detectar e resumir (*summarize*) o que está nos dados, e para isso já existem boas ferramentas que foram desenvolvidas com esse intuito, como a regressão linear (Varian, 2014). Por outro lado, os algoritmos de *machine learning* geralmente funcionam de uma forma diferente, com um foco maior na tarefa de previsão da variável dependente, e não tanto nessas outras relações. Assim, embora seja verdade que esses algoritmos venham tendo êxito em seu propósito, eles também apresentam uma espécie de *black-box* em relação à sua interpretabilidade, isto é, não fica exatamente muito claro como o algoritmo chegou naquela decisão, qual função aplicou aos dados e como as variáveis predictoras afetam a variável target, o que por vezes não seria interessante para econométricos. James et al. (2013) apresentam um *trade-off* entre interpretabilidade e flexibilidade/acurácia no sentido em que as técnicas mais usadas por economistas como os Mínimos Quadrados Ordinários (MQO) ou métodos paramétricos em geral mostram pouca flexibilidade, mas são muito fáceis de serem interpretados, enquanto as técnicas de *machine learning* (métodos não paramétricos) usualmente têm grande poder de acurácia, apesar de serem muito mais difíceis de serem interpretados.

Desse modo, a escolha de qual técnica usar passará necessariamente pelo o que o pesquisador deseja com seu trabalho. Ainda assim, nem sempre o método mais flexível fará as melhores previsões. Já se sabe pela teoria econômica que alguns modelos com características simples não ficam tão atrás de técnicas mais rebuscadas, chegando até mesmo a mostrarem resultados superiores em certos cenários. Fazer esse tipo de investigação para a inflação brasileira será um dos objetivos desta dissertação.

## Os tipos de aprendizado de máquina

Existem dois tipos principais de aprendizado de máquina: supervisionado e não supervisionado. A diferença entre eles é a existência ou não de uma variável dependente/target (Y). Ou seja, no caso mais comum, que é o de aprendizado supervisionado, será usado um grupo de variáveis preditoras (*features*) para modelar a variável dependente, e então generalizar para conseguir fazer previsões de Y para novas observações (James et al, 2013). A maioria dos métodos mais conhecidos de *machine learning* se enquadra nesse primeiro grupo, assim como os que foram usados nesta dissertação, dado que a previsão da inflação é uma tarefa supervisionada. Alguns desses métodos são a própria regressão linear, o *random forest* e as redes neurais. Já para o tipo de aprendizado não supervisionado, não existe a variável target (Y). Nesse caso, o algoritmo em questão tentará encontrar algum tipo de padrão nos dados, e criará grupos com base nesse padrão, geralmente em tarefas do tipo *clustering* (Chakraborty & Joseph, 2017). Assim, na hora de fazer previsão, novas observações serão adicionadas em cada um desses grupos.

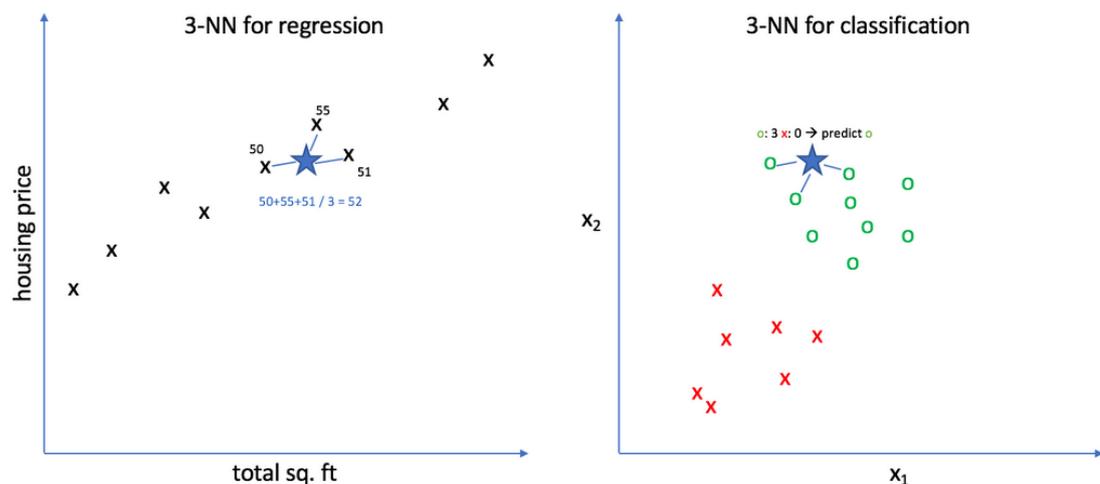
Então, uma vez selecionado o aprendizado supervisionado, é comum que se faça uma distinção entre os tipos de problemas encontrados pelo pesquisador, que podem ser de dois tipos: classificação ou regressão. Na realidade, como explicam James et al. (2013), existe uma divisão prévia das variáveis entre qualitativas e quantitativas. As quantitativas são números, que podem representar preços, rendimentos, anos de experiência etc. Por sua vez, as qualitativas têm a ver com a classe das variáveis, como o gênero de uma pessoa, a espécie de um animal ou planta, se um evento irá ocorrer ou não, entre outros. Dessa maneira, os problemas que envolvem prever uma variável do tipo quantitativa são chamados de problemas de regressão (que não tem o mesmo significado de econometria, pois nesta o termo significa um tipo de modelo, enquanto em machine learning diz respeito a um tipo de classe – Chakraborty & Joseph, 2017), ao passo que naqueles problemas que a variável target é qualitativa, estamos diante de um problema do tipo de classificação. Como o objetivo desta dissertação é fazer a previsão de um índice de inflação (variável contínua, numérica), será realizada uma tarefa de regressão, ainda que a maioria dos métodos realize os dois tipos.

### 2.2.2.1 k-Nearest Neighbors (KNN)

O método dos KNN é um dos algoritmos de aprendizado supervisionado mais simples em *machine learning*. Sua premissa básica é atribuir a uma nova observação o valor ou classe de acordo com um número K de vizinhos próximos a ela.

Dessa forma, o primeiro passo exige que se calcule a distância entre uma nova observação que será instanciada e as demais variáveis preditoras (*features*) em um plano. Após isso, será determinado o número K de vizinhos mais próximos ao lugar onde essa nova observação será instanciada, e a partir disso poderá se descobrir (prever) a qual grupo ela pertence (Chakraborty & Joseph, 2017). No caso de um problema de classificação, a classe será definida de acordo com o voto majoritário (embora o termo majoritário não se aplique tão bem nesse caso, sendo melhor substituído por ‘pluralitário’ – Raschka, 2018, isso não interfere no entendimento) das outras classes que estão dentro daquela distância; ou seja, dado o número K de vizinhos, se essa nova observação for instanciada em um local que tenha mais observações de uma classe A do que uma classe B, a instância será classificada como classe A. Nos problemas de regressão, como as *features* são variáveis numéricas, a forma de se atribuir um valor para a nova observação será por meio da média aritmética entre os K-vizinhos daquela área. A figura abaixo torna a tarefa de entender o funcionamento do algoritmo muito mais fácil.

**Figura 3 – Previsão de KNN para k = 3. Fonte: Jeremy Jordan<sup>7</sup>.**



<sup>7</sup> Uma breve explicação de KNN com script em Python. Disponível em: <https://www.jeremyjordan.me/k-nearest-neighbors/>

No painel da esquerda da figura, em um problema de regressão, a nova observação (representada por uma estrela azul) recebe o valor da média aritmética dos 3 vizinhos mais próximos do local onde ela foi instanciada. Enquanto isso, no painel da direita, que é um problema de classificação, ela é classificada como sendo da classe verde.

Ainda que existam algumas formas de se calcular a distância entre os pontos, e que esse cálculo dependa bastante do contexto do problema, não existe exatamente uma melhor forma (Raschka, 2018). Mas as distâncias mais conhecidas e usadas são 3: 1) distância euclidiana; 2) distância Manhattan; e 3) distância Minkowski. A euclidiana costuma ser a mais usada entre elas, sobretudo para os problemas de regressão. A distância Minkowski é a generalização das distâncias euclidiana e Manhattan, e tem a seguinte representação:

$$d(x, y) = (\sum_{i=1}^n |x_i - y_i|)^{\frac{1}{p}} \quad (10)$$

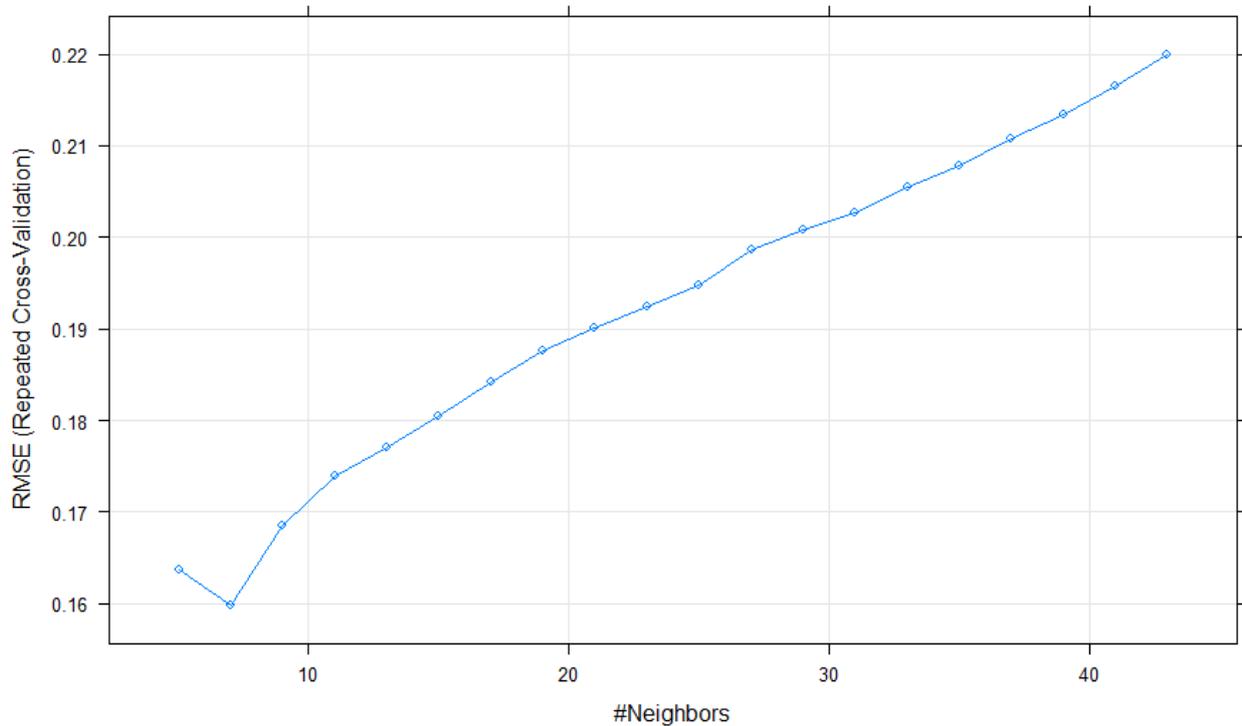
Quando o parâmetro ‘p’ é igual a 1, temos a distância Manhattan. Quando ele é igual a 2, a distância euclidiana. Outros valores de ‘p’ permitem que sejam criadas outras medidas de distância<sup>8</sup>.

Em relação à escolha do número ótimo de K para determinada tarefa, há uma espécie de trade-off entre viés e variância, no sentido em que valores menores de K tendem a ter uma alta variância, baixo viés e um alto grau de complexidade. Uma configuração desse tipo (alta variância e baixo viés) pode gerar o problema de *overfitting*, quando acaba se gerando um número elevado de parâmetros. Por outro lado, valores de K mais altos costumam ter uma baixa variância, alto viés e baixa complexidade. Manualmente, talvez seja um pouco trabalhoso escolher um bom número para K, além de que se deve evitar números pares em tarefas de classificação para que não ocorram empates. Contudo, na prática, em machine learning, o processo de achar o melhor número de K é chamado de tunagem (*tuning*) dos hiperparâmetros ou model selection, em uma etapa chamada de validação cruzada (Raschka, 2018). Felizmente, a validação cruzada é feita pelo próprio software, mas o cientista precisa imputar um conjunto de valores para os parâmetros que faça sentido, e então cada um deles será testado e o melhor será escolhido de acordo com alguma métrica.

**Figura 4 – Número de K-vizinhos escolhido no software Rstudio por validação cruzada.**

---

<sup>8</sup> Outra breve explicação sobre KNN em: <https://www.ibm.com/topics/knn>



Como pode ser observado na figura 4, especificamente para esta dissertação, o valor de  $K = 7$  foi escolhido (por ter menor RMSE) para o número de vizinhos na etapa de validação cruzada fixando o máximo de 20 valores intervalados para serem testados. Em problemas de regressão, o Rstudio já escolhe automaticamente o valor de  $K$  que possui o menor RMSE, mas é possível modificar para usar como métrica o menor MAE ou maior  $R^2$ . Para problemas de classificação, o software já escolhe automaticamente aquele  $K$  que apresenta a maior acurácia, isto é, aquele  $K$  que gera um maior índice de acerto na etapa de treinamento do modelo.

### 2.2.2.2 Support Vector Machines (SVM)

Os *support vector machines* correspondem a um outro tipo de técnica de *machine learning* não-paramétrica que pode realizar tarefas de classificação e regressão. Para efeito de simplificação, será dada uma explicação sobre este algoritmo para um problema de classificação com apenas duas classes, assim como fazem Athey & Imbens (2019) e Cortes & Vapnik (1995). Dessa maneira, o primeiro passo é definir um hiperplano. Supondo que ele seja linear (e que as classes sejam linearmente separáveis), define-se  $X_i$  como um vetor de *features* (ou variáveis preditoras), e  $\omega$  como um vetor de pesos, com  $N$  observações para uma variável

de saída ( $y$ ) que pode ter dois resultados: -1 e 1. Assim, o hiperplano pode ser representado com a seguinte função  $y = \omega^T x + b$ . Como as classes são linearmente separáveis, podemos dizer que as inequações abaixo são válidas para todos os dados de treino (Cortes & Vapnik, 1995):

$$\begin{aligned} \omega^T x + b &\geq 1 && \text{se } y_i = 1 \\ \omega^T x + b &\leq -1 && \text{se } y_i = -1 \end{aligned} \quad (11)$$

E se cada uma das equações se igualar a -1 e 1, teremos duas retas paralelas (chamadas de margens) no espaço das *features* (nome dado a um plano cartesiano com as duas classes em questão). Dessa forma, o primeiro objetivo do algoritmo é encontrar o hiperplano ótimo que passa entre essas duas retas em (11). Esse hiperplano terá a seguinte configuração:

$$\omega^T x + b = 0 \quad (12)$$

Então, uma vez que os dados são linearmente separáveis por um hiperplano ótimo, o próximo passo é encontrar a distância máxima entre esse hiperplano e as margens, que pode ser dada pela seguinte equação:

$$k = \frac{2}{\|\omega\|} \quad (13)$$

Onde  $\|\omega\|$  é a norma do vetor  $\omega$ . O algoritmo então tentará maximizar essa distância entre o hiperplano e as margens, que é a mesma coisa que minimizar  $\|\omega\|$  ou  $\frac{1}{2} \|\omega\|^2$ . Os pontos que estão na mesma distância entre as margens e o hiperplano são chamados de vetores de suporte (*support vectors*).

Antes de prosseguir, se multiplicarmos  $y_i$  nos dois lados das inequações em (11), obteremos o seguinte (Cortes & Vapnik, 1995):

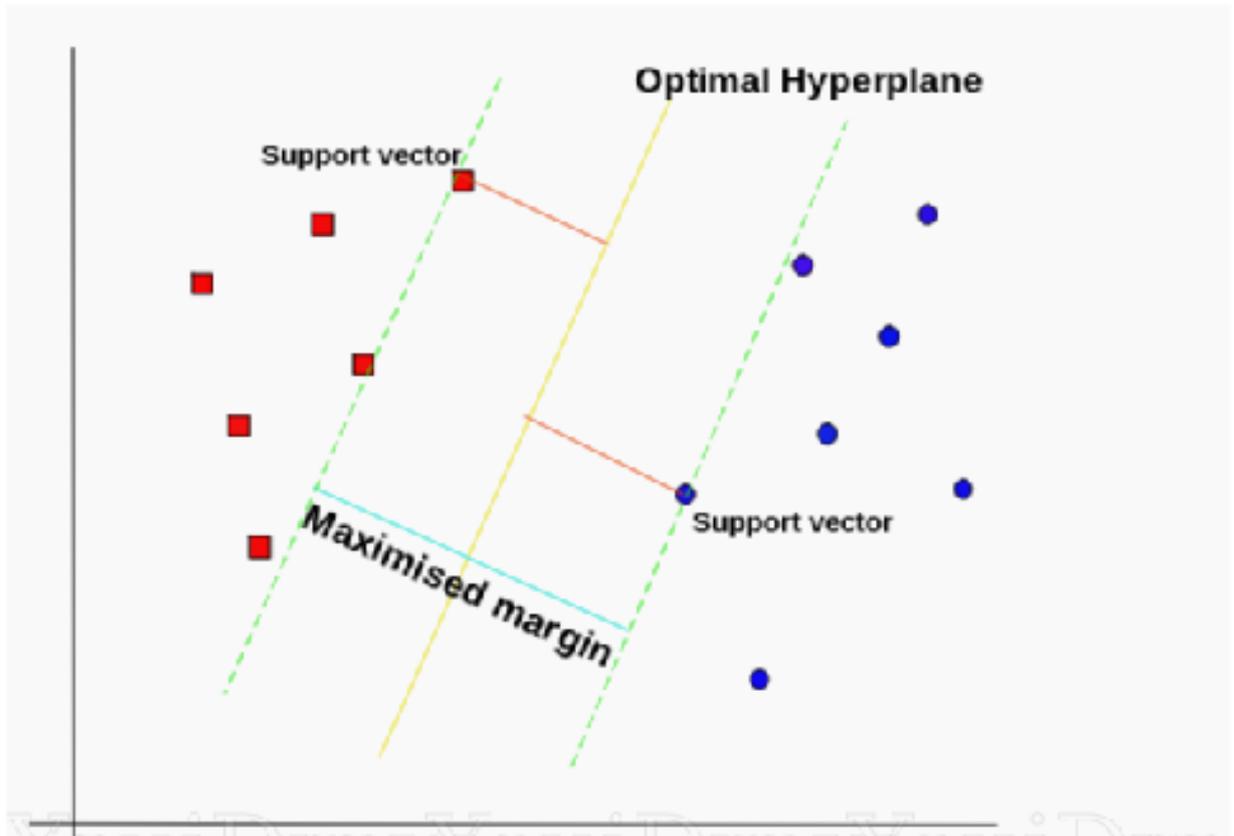
$$y_i(\omega^T x_i + b) \geq 1 \quad (14)$$

Por fim, o problema de otimização para um caso de SVM em que os dados são linearmente separáveis fica:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 \quad \text{sujeito a } y_i(\omega^T x_i + b) \geq 1, \quad \forall i = 1, \dots, N \quad (15)$$

A figura 5 ajuda a ilustrar o que foi exposto acima.

**Figura 5 – Representação de um SVM linearmente separável. Fonte: Towards Data Science**



Porém, na maioria das vezes, os dados não serão linearmente separáveis, e terão de ser projetados em um plano com dimensões maiores, que vão além do 2D. Para isso, um determinado tipo de transformação deverá ser aplicada aos dados. Embora essa tarefa seja difícil, existe um facilitador que encurta o tempo e o custo computacional; tal facilitador se chama de “Truque” do Kernel (ou Kernel Trick). Antes de introduzir esse Truque, é melhor que se apresente o problema de otimização (primal) da equação (15) como um problema dual com o uso do Lagrangeano. Portanto, reescrevendo (15):

$$L = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^N \alpha_i [y_i(\omega^T x_i - b) - 1] \quad (16)$$

É interessante olhar o problema nessa formulação porque, ao resolvê-lo, a maximização dependerá somente do produto interno dos vetores  $x_i$  e  $x_j$  ( $x_j$  só aparece ao resolver a otimização), o que simplifica o problema em muitos casos já que não será preciso passar por todas as observações, apenas aquelas em que o lagrangeano ( $\alpha_i$ ) será maior ou igual a zero, isto é, apenas nos vetores de suporte.

Após isso, ao nos depararmos com dados que não são linearmente separáveis, deverá existir uma função  $\varphi(x)$ , na qual 'x' é um vetor, que transforma os dados para um espaço de maior dimensão. Como a otimização depende apenas do produto interno de  $x_i$  e  $x_j$ , não é preciso que essa função seja especificada (Athey & Imbens, 2019). O que será feito, na realidade, é a otimização do lagrangeano usando a transformação de  $\varphi(x_i) \cdot \varphi(x_j)$ , em uma função chamada de Kernel. Ela pode ser descrita em termos de um produto interno  $\langle \cdot \rangle$  como:

$$K(x_i, x_j) = \langle \varphi(x_i) \cdot \varphi(x_j) \rangle \quad (17)$$

Ou seja, o “Kernel Trick” realizará o produto interno dos vetores em outra dimensão, não sendo necessário saber qual tipo de transformação será feita para mudar de dimensão, até mesmo porque saber exatamente qual seria essa função costuma ser uma tarefa relativamente difícil (Chakraborty & Joseph, 2017). Duas das funções kernels mais conhecidas são a polinomial e a gaussiana (radial basis function). Como a maioria dos problemas que envolvem SVM não serão linearmente separáveis, é importante entender a forma que ele funciona mais comumente.

Na prática, ao usar o Rstudio para a tarefa de prever o IPCA, o software achou a melhor configuração como sendo aquela com 70 vetores de suporte e uma função de kernel do tipo gaussiana (radial). Ou seja, ao realizar o truque do kernel e otimizar a função de perda, essa especificação achou o menor RMSE dentro dos dados de treino. Ainda que aqui seja um problema de regressão, a noção é basicamente a mesma da que foi apresentada acima, com algumas poucas mudanças e pode ser vista melhor em Chakraborty & Joseph (2017). O SVM é mais robusto que o KNN, tendo melhores resultados em bases de dados mais complexas, e tende a gerar menos overfitting, apesar do alto custo computacional.

### 2.2.2.3 LightGBM

O algoritmo de LightGBM está incluído em um tipo de técnica chamada de *ensemble learning*. Em machine learning, o termo *ensemble* significa a junção de vários modelos com o objetivo de se atingir um resultado preditivo melhor. Embora na econometria esse tipo de agregação de modelos não seja algo novo, Athey & Imbens (2019) pontuam que a diferença é que geralmente no caso da econometria os modelos combinados costumam ser parecidos, mas com parâmetros variados, enquanto que em ML eles podem ser bastante diferentes e os pesos para fazer uma média entre os modelos surja a partir de um processo um pouco distinto. Dois dos grupos que fazem ensemble são o bagging e o boosting, que buscarão agrupar vários *weak learners* (estimadores que possuem um erro de previsão relativamente alto) em um *strong learner* (que possui erro de previsão menor). A técnica de bagging consiste em fazer a média das previsões de um grupo de estimadores independentes com o propósito de diminuir a variância e consequentemente melhorar as previsões.

Por outro lado, a técnica de boosting irá fazer repetidas previsões nos dados de treino, e aquelas que tiverem um erro maior, terão um peso proporcionalmente maior em cada uma das novas previsões (Varian, 2014). Dessa forma, cada um dos próximos estimadores tentará prever os resíduos do modelo imediatamente anterior e ir criando uma árvore de decisão para cada um deles, sendo que essas árvores têm um formato um pouco diferente das demais, com apenas um nó inicial e duas folhas (*decision tree stumps*), que por definição são *weak learners*. Neste caso, cada nova árvore construída dependerá da anterior porque os estimadores serão construídos sequencialmente com a intenção de minimizar o erro em cada uma das iterações.

Uma evolução da técnica de boosting é a de Gradient Boosting, que usará o negativo do vetor gradiente para tentar minimizar o erro de uma função de perda, fazendo com que cada nova árvore adicionada se aproxime cada vez mais da variável dependente<sup>9</sup>. Essa minimização não será feita de forma imediata como em um problema de MQO, por exemplo, mas sim fazendo pequenas mudanças ao calcular o gradiente em direção ao ponto em que a perda é mínima (na técnica chamada de gradiente descendente; no caso de uma função quadrática, por exemplo, seriam testados valores para o intercepto e a inclinação paulatinamente). A diferença é que poderão existir situações em que não será possível chegar no ponto da derivada igual a zero, então o algoritmo chegaria em um ponto próximo o suficiente. Aqui, as árvores que serão construídas não terão mais a característica de apenas um nó e duas folhas (*stumps*), podendo ser maiores, mas sempre serão multiplicadas por um valor (*learning rate*) para evitar que ocorra

---

<sup>9</sup> Uma explicação mais detalhada sobre Gradient Boosting pode ser lida em Mason et al. (1999)

overfitting (ou seja, a learning rate impedirá que o valor predito seja exatamente igual à variável dependente nos dados de treino, senão o modelo ficaria mal treinado).

Isso posto, um dos algoritmos mais conhecidos de machine learning é o Extreme Gradient Boosting (XGBoost), que representa uma melhoria em relação ao Gradient Boosting. Por conta de realizar otimizações algorítmicas, ele tem uma escalabilidade maior do que o Gradient Boosting normal, e conseqüentemente consegue explorar melhor as configurações de hardware de um computador, realizando mais simulações em um tempo menor. Basicamente, ao usar computação paralela (isto é, seleciona a base de dados, que geralmente é grande, divide em blocos menores, e trabalha em cada uma delas separadamente, mas ao mesmo tempo), ele é mais rápido e permite que o modelo seja treinado de uma forma melhor. Também utiliza outras técnicas como o tratamento de dados esparsos, que geralmente costumam ser valores faltantes (NA's) ou dummies (que são chamadas de one-hot-encoding em ciência de dados), sem que se precise transformar esses dados, pois o próprio algoritmo tem um procedimento default para usar. Em resumo, o XGBoost é mais eficiente porque ele é computacionalmente mais robusto. E essa eficiência é corroborada em diversas competições de ciência de dados nas quais as soluções vencedoras usam esse algoritmo, além do crescimento em problemas reais de empresas (Chen & Guestrin, 2016). Porém, uma vez que Araújo & Gaglianone (2023) já o aplicaram, e esta dissertação tem como um dos objetivos complementar aquele artigo, ele não será usado aqui.

Mais recentemente, surgiu um novo método bastante similar ao XGBoost, chamado de Light Gradient Boosting Model (LightGBM). A ideia principal desse algoritmo é tentar aumentar a escalabilidade e eficiência em grandes bases de dados. Como normalmente nos casos de Gradient Boosting existe uma perda de tempo para escolher a melhor configuração de divisão (split) dos dados para criar as árvores de decisão (quando o volume desses dados é muito grande (big data)), isso pode eventualmente prejudicar o tempo de processamento e tornar essa implementação custosa. Então, são duas as inovações trazidas pelo LightGBM que buscam melhorar a rapidez no processamento, quais sejam: Gradient-Based One-Side Sampling (GOSS) e o Exclusive Feature Bundling (EFB). A primeira delas é focar nas instâncias com maior gradiente (em módulo) para fazer a amostragem nos dados, e escolher os melhores pontos de split da árvore de decisão. Dado que um gradiente menor (em módulo) implica que aquela instância está melhor treinada e, portanto, possui um erro de previsão mais baixo, o melhor a se fazer então é se empenhar naquelas que têm um gradiente mais alto, e que trazem um ganho de informação maior para os dados. Dessa forma, todos os pontos de gradiente maior são mantidos na hora de fazer a amostragem, e apenas algumas instâncias com gradiente menor são

escolhidas de forma aleatória (elas não são excluídas por completo para não mudar a distribuição dos dados). Os autores afirmam que esse tipo de amostragem nos dados é mais eficiente que uma amostragem aleatória total, além de reduzir o esforço computacional (Ke et al. 2017).

Já o EFB trabalhará no fato de que, em um grande conjunto de dados, grande parte deles será esparso. Assim, essa técnica agrupará as variáveis preditoras (*features*) que são mutuamente exclusivas, diminuindo o número de variáveis preditoras que serão treinadas, isto é, em vez de fazer o treinamento de todas as *features*, serão treinados os grupos (*bundles*) que foram criados, de maneira que eles são menores em quantidade do que as *features*. É importante lembrar que esses algoritmos são construídos para trabalhar preferencialmente com grandes bases de dados, que podem conter milhares de *features*. Nesses casos, a complexidade do processo de treinamento e o custo computacional é bastante reduzido. O LightGBM é o modelo de boosting mais rápido em termos de hardware, sem perder a acurácia das previsões, superando até mesmo o XGBoost (Ke et al. 2017).

Assim, nesta dissertação, no que tange aos modelos de machine learning, utilizaremos os modelos KNN, SVM e LightGBM. Os resultados são apresentados no próximo capítulo.

## 2.3 Critérios de avaliação

Há vários critérios de avaliação de modelos de previsão. Para decidir qual dos modelos apresentados acima mostrou-se o mais eficaz, utilizaremos três critérios: erro quadrático médio (MSE, para melhor conveniência), raiz do erro quadrático médio (RMSE) e o teste de Diebold-Mariano (1995).

### 2.3.1 Erro quadrático médio (MSE)

O erro quadrático médio (MSE) possui a seguinte formulação:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (18)$$

Onde  $y_i$  é a variável dependente observada e  $\hat{y}_i$  é a sua previsão.

Nota-se que as unidades do MSE são o quadrado das unidades originais dos dados, tornando a interpretação dos valores mais difícil. Valores do MSE são usados durante o treinamento em problemas de otimização onde a diferenciação é importante (como no caso de método do gradiente).

### 2.3.2 Raiz do erro quadrático médio (RMSE)

A raiz do erro quadrático médio (RMSE) possui a seguinte formulação:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

A unidade da raiz do erro quadrático médio (RMSE) é a mesma da variável sendo predita, o que torna a sua interpretação mais simples.

Para ambos, o MSE e o RMSE, quanto menor o valor, melhor a acuracidade da previsão. No caso do RMSE, em particular, o valor absoluto do indicador deve ser analisado no contexto da amplitude de valores apresentada pelos dados. Por exemplo, um  $RMSE = 4$  será considerado alto se os dados originais variam entre 0 e 10, mas será considerado baixo se variam entre 0 e 5000. Por isso, muitas vezes, o RMSE é expresso em termos relativos (em relação à média dos dados), como a razão  $RMSE/média$  dos dados.

Quando se está comparando diversas estimações do ponto de vista da acuracidade, no entanto, basta que se compare os resultados de cada métrica entre si, isto é, dentre dois modelos de previsão, o melhor será aquele que de menor MSE (ou RMSE).

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

O referido teste visa comparar se duas técnicas de previsão distintas são significativamente diferentes de fato, do ponto de vista estatístico. Para tanto, supondo que  $\hat{y}_{it}$

e  $\hat{y}_{jt}$  são as previsões de  $y_t$ , com erros de previsão  $e_{it}$  e  $e_{jt}$ , podemos definir a função de perda  $g(\cdot)$  como sendo uma função dos erros, ou seja,

$$g(y_t, \hat{y}_{it}) = g(e_{it}) \text{ e } g(y_t, \hat{y}_{jt}) = g(e_{jt}) \quad (20)$$

Então, define-se o diferencial dessa perda como:

$$d_t = g(e_{it}) - g(e_{jt}) \quad (21)$$

Dado que o teste busca comparar se as duas previsões têm a mesma acurácia, as hipóteses são apresentadas como:

$$\begin{aligned} H_0: E(g(e_{it})) &= E(g(e_{jt})) \quad \text{ou} \quad E(d_t) = 0 \\ H_1: E(g(e_{it})) &\neq E(g(e_{jt})) \quad \text{ou} \quad E(d_t) \neq 0 \end{aligned} \quad (22)$$

Portanto, a hipótese nula é de que as duas possuem acurácias estatisticamente iguais, enquanto que a hipótese alternativa é que são acurácias diferentes. Assumindo que o diferencial da função de perda é estacionário em covariância e de curta memória, Diebold & Mariano (1995) definem a média assintótica dessa função:

$$\bar{d} = \frac{1}{T} \sum_{t=1}^T [g(e_{it}) - g(e_{jt})] \quad (23)$$

Ademais, também deve se introduzir  $f_d(0)$  como sendo a densidade espectral de  $d_t$  na frequência 0:

$$f_d(0) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau) \quad (24)$$

Sendo que  $\gamma_d(\tau)$  é a auto covariância de  $d_t$  no lag  $\tau$ . Como se assumiu que  $d_t$  é estacionário em covariância e de curta memória, pode-se definir que a estatística abaixo converge assintoticamente em distribuição para uma normal:

$$\sqrt{T}(\bar{d} - \mu) \rightarrow N(0, 2\pi f_d(0)) \quad (25)$$

É possível reescrever (25) como convergindo para uma normal padrão:

$$\frac{\bar{d} - \mu}{\sqrt{\frac{2\pi f_d(0)}{T}}} \rightarrow N(0,1) \quad (26)$$

Sob  $H_0$ , com  $\mu = 0$ , reescrevemos (26):

$$\frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}} \rightarrow N(0,1) \quad (27)$$

Dessa forma, quando consideramos o horizonte de previsão de  $\tau = 1$ , a estatística de teste de Diebold-Mariano é a mesma de (27):

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}} \quad (28)$$

Porém, ao considerar  $\tau \geq 1$ , a fórmula apresentada em (24) será expandida para sua forma final, representada como<sup>10</sup>:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\gamma_d(0) + 2 \sum_{k=1}^{\tau-1} \gamma_d(k)}{T}}} \quad (29)$$

---

<sup>10</sup> Essa derivação não será feita aqui, mas o leitor pode conferir o artigo de Diebold-Mariano (1995) para mais detalhes.

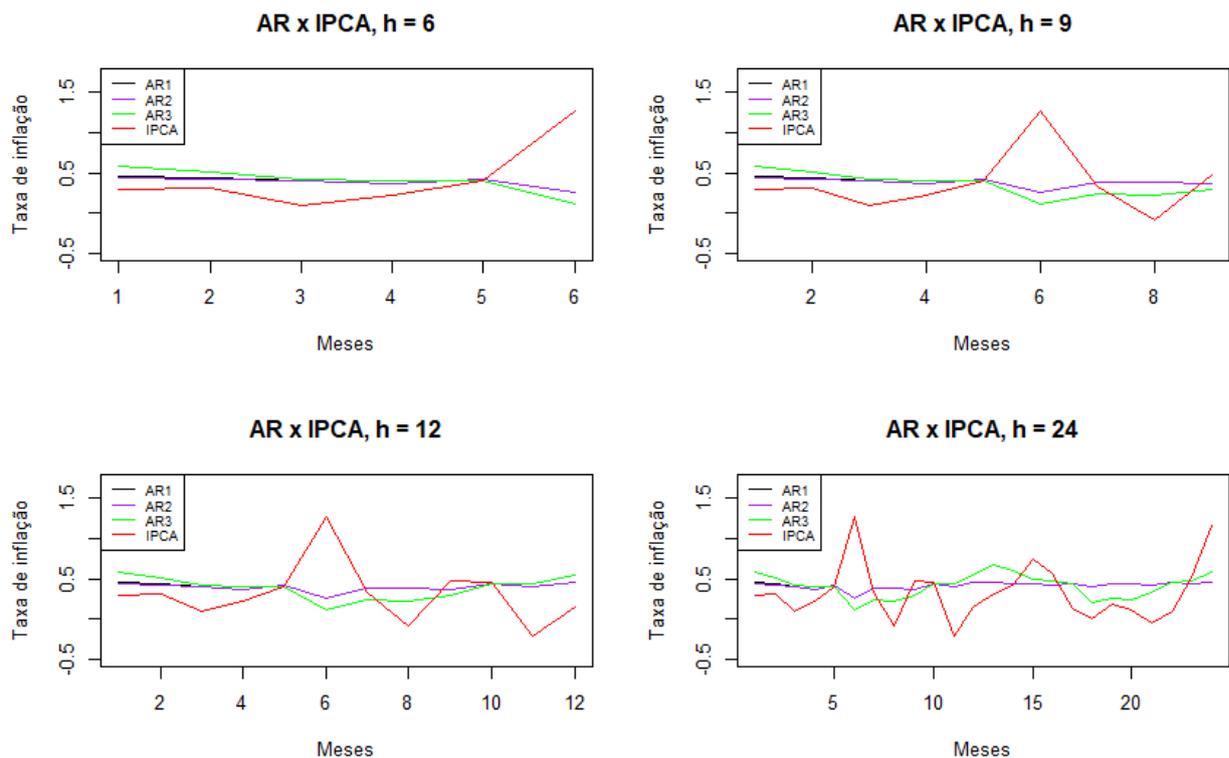
## CAPÍTULO III – RESULTADOS

Finalmente, este capítulo apresentará os resultados obtidos por meio das simulações no software. Em um primeiro momento, cada um dos métodos será exposto apenas em comparação ao IPCA, enquanto que, posteriormente, todos aparecerão conjuntamente com o índice. Ademais, haverá também uma tabela com os critérios de avaliação e depois o teste Diebold & Mariano para verificar se as eventuais diferenças de previsão são estatisticamente significativas.

### 3.1 Modelos

#### 3.1.1 Modelo autorregressivo

Figura 6 – AR x IPCA



Conforme foi visto no capítulo 2, a série do IPCA para o caso brasileiro apresenta um alto grau de autocorrelação. Mais especificamente, há autocorrelação até o *lag* 3. Dessa forma, também foram incluídos os modelos AR (2) e AR (3). Ademais, a presença de sazonalidade na série foi tratada no software para cada um dos modelos individualmente (usando o parâmetro

*seasonal*). No gráfico, as linhas dos três modelos exibem uma trajetória bastante similar, dando a impressão que eles possuem desempenhos semelhantes. Para ter certeza ou não dessa impressão, a melhor forma é olhar as métricas de desempenho, que são o MSE e o RMSE. A tabela 2 abaixo faz tal tarefa.

**Tabela 2 – Erros de previsão modelo autorregressivo**

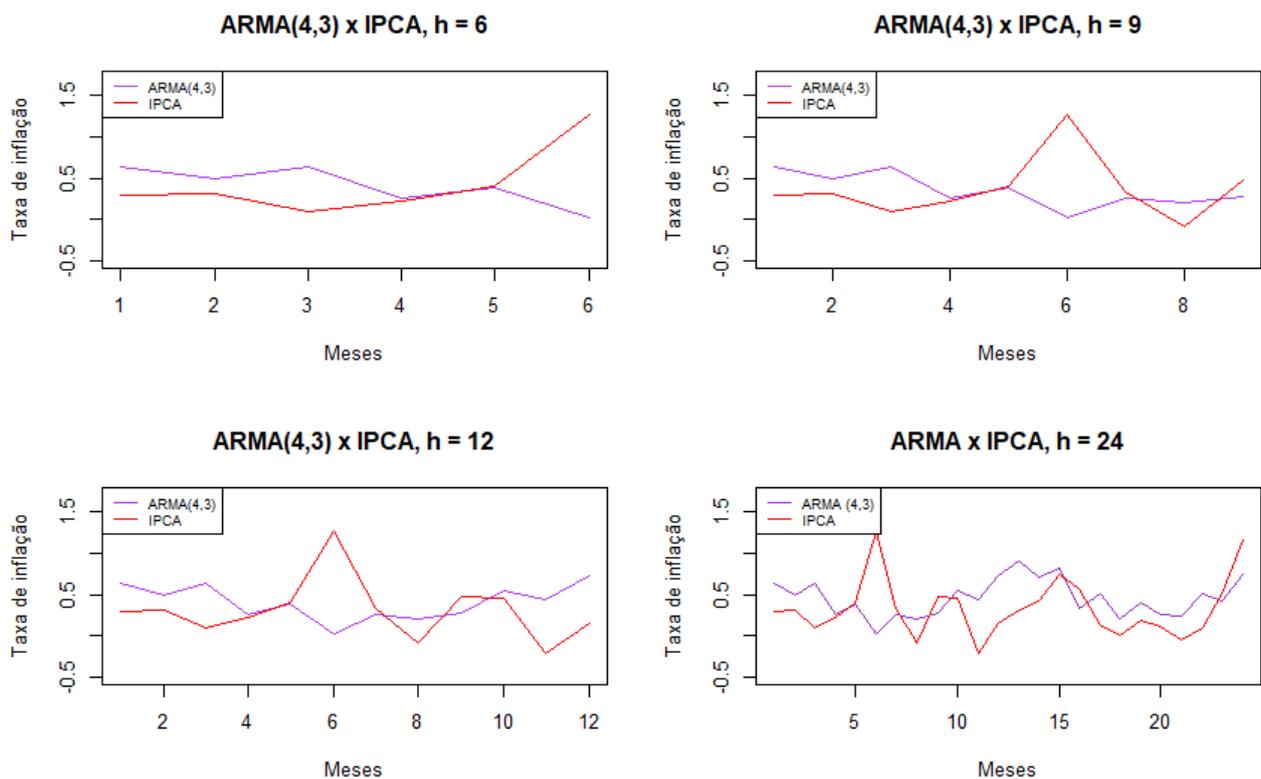
|             | Erros de previsão |           |           |           |           |           |           |           |
|-------------|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|             | 6 meses           |           | 9 meses   |           | 12 meses  |           | 24 meses  |           |
|             | MSE               | RMSE      | MSE       | RMSE      | MSE       | RMSE      | MSE       | RMSE      |
| <b>AR 1</b> | 0.1926843         | 0.4389582 | 0.1546484 | 0.3932536 | 0.1547431 | 0.393374  | 0.1349035 | 0.3672921 |
| <b>AR 2</b> | 0.1929717         | 0.4392854 | 0.1547639 | 0.3934004 | 0.1547646 | 0.3934014 | 0.1348266 | 0.3671875 |
| <b>AR 3</b> | 0.2613629         | 0.5112366 | 0.1901732 | 0.4360885 | 0.1910825 | 0.4371299 | 0.1355853 | 0.3682191 |

Embora os erros de previsão do AR (1) e do AR (2) sejam de uma magnitude muito parecida entre eles, ao olharmos com atenção, podemos perceber que os erros do modelo AR1 são ligeiramente menores, para quase todos os horizontes de previsão. A questão que se coloca, portanto, nesse primeiro momento, é se essa diferença de acurácia é, de fato, estatisticamente significativa. Para respondê-la, fizemos uso do teste de Diebold-Mariano para um nível de significância de 5% (no teste foram usados os resíduos puros, sem transformação, tal como é exemplificado na documentação do método *dm.test* no R). O resultado do teste mostrou que, para todos os horizontes, os erros de previsão são estatisticamente iguais, com p-valores relativamente altos. Ou seja, ainda que o AR (1) tenha erros menores, eles não são diferentes do AR (2), e nem mesmo do AR (3), em termos estatísticos. Contudo, como desejamos usar só um dos modelos autorregressivos para comparar com os outros que virão a seguir, é mais conveniente escolher o AR (1), por ser o mais parcimonioso.

### 3.1.2 Modelo ARMA (4,3)

Conforme ressaltado no capítulo anterior, o intuito de adicionar o modelo ARMA (4,3) é o fato de o Banco Central do Brasil ter usado o modelo com essa especificação no período que compreende 2001 até 2018, isto é, toda a esfera que essa dissertação usou. Na verdade, um período maior.

**Figura 7 – ARMA x IPCA**



O pico do IPCA no mês 6 é um movimento relevante que impacta a previsão fora da amostra para todos os métodos, e que graficamente contrasta com o desempenho dos modelos. No caso do ARMA, a performance do modelo só melhora minimamente a partir do período de 9 meses. Isso é corroborado pelos erros de previsão que diminuem comparativamente entre  $h = 9$

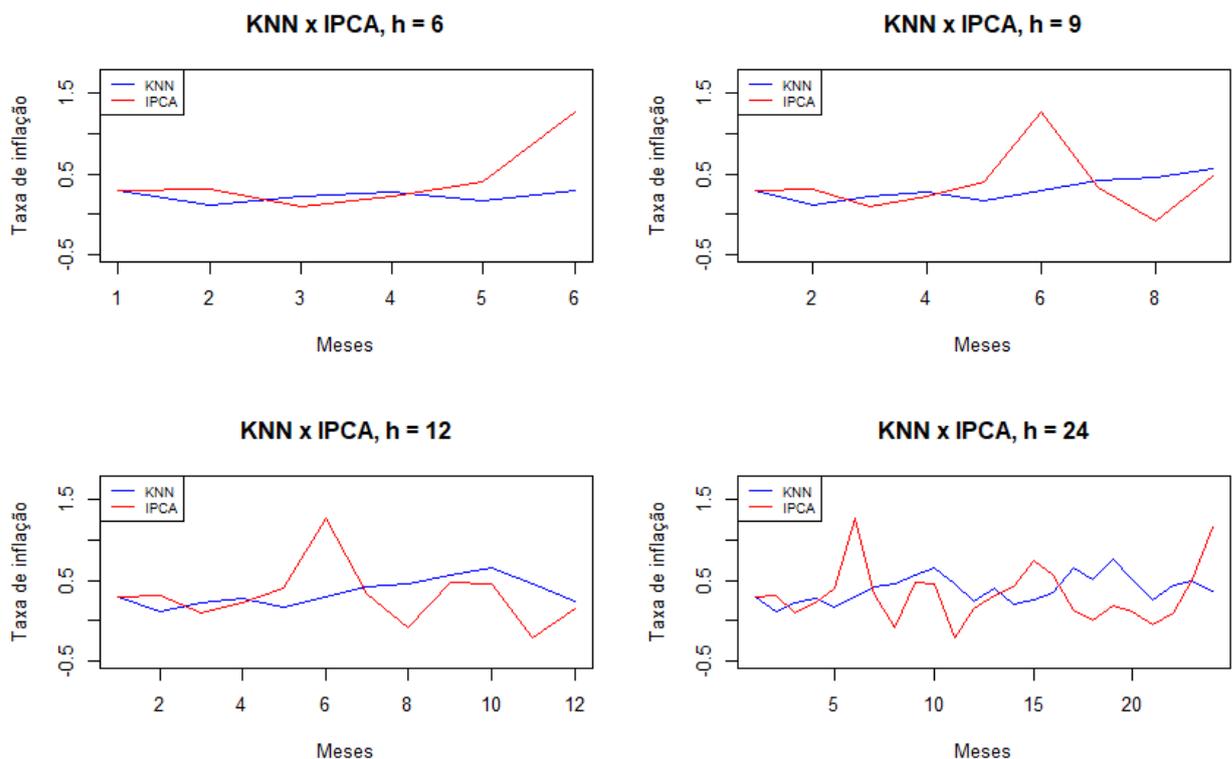
(MSE de 0.235 e RMSE de 0.485) e  $h = 24$  (MSE de 0.167 e RMSE de 0.409), ainda que graficamente não pareça tão claro.

### 3.1.3 Modelo KNN

A partir de agora, serão apresentados os modelos de machine learning. Como já dito no capítulo 2, o *k-Nearest Neighbour* é um dos modelos mais básicos de aprendizado de máquina, mas que pode apresentar bons resultados, apesar de suas limitações.

Várias especificações do modelo KNN foram estimadas, utilizando-se diferentes pacotes do Rstudio, com resultados bastante parecidos. A utilização de diferentes configurações de validação cruzada também não apresentou diferenças muito grandes. Os resultados mostrados aqui são do modelo KNN estimado com o pacote *caret* e um número de 7 vizinhos.

**Figura 8 – KNN x IPCA**



Novamente ao se observar o gráfico, o pico do índice no mês 6 chama a atenção pelo fato de o modelo não ter conseguido acompanhar. Por outro lado, mesmo que não fique tão conclusivo pelo gráfico, os erros de previsão para o KNN são menores do que os do ARMA (4,3) para todos os períodos. Em relação ao AR (1), apenas para  $h = 6$  e  $h = 9$ . Isso poderá ser

visto mais claramente na tabela 6, onde serão comparados os erros de previsão de todos os modelos. A tabela 3 traz apenas os erros de previsão do KNN:

**Tabela 3 – Erros de previsão KNN**

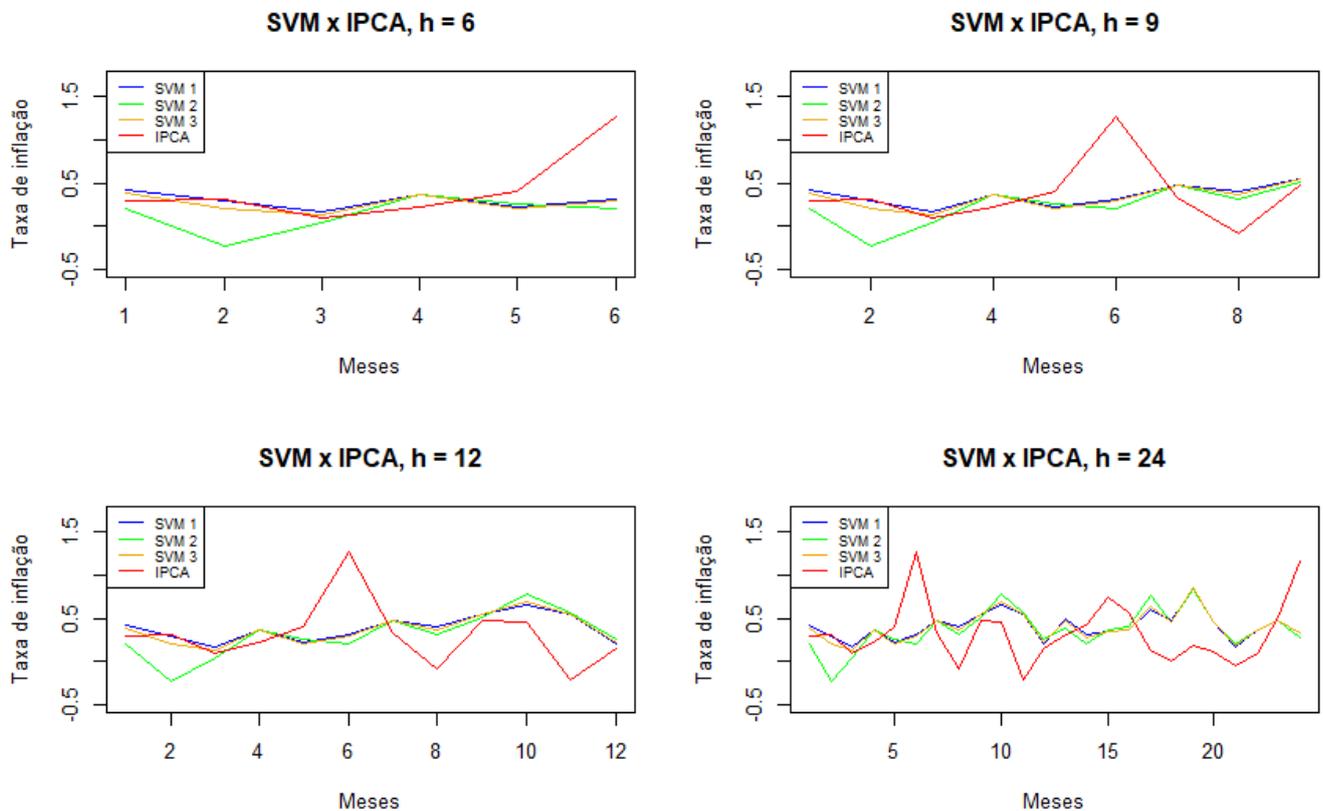
|            | Erros de previsão |          |           |           |           |           |           |           |
|------------|-------------------|----------|-----------|-----------|-----------|-----------|-----------|-----------|
|            | 6 meses           |          | 9 meses   |           | 12 meses  |           | 24 meses  |           |
|            | MSE               | RMSE     | MSE       | RMSE      | MSE       | RMSE      | MSE       | RMSE      |
| <b>KNN</b> | 0.1758133         | 0.419301 | 0.1522636 | 0.3902096 | 0.1559303 | 0.3948802 | 0.1684778 | 0.4104605 |

Por um lado, comparados aos modelos de séries de tempo, exhibe uma certa evolução para os horizontes mais curtos (até 9 meses), embora não tenha conseguido bater o AR (1) a partir de 12 meses. Por outro lado, será interessante notar como o KNN performará frente aos dois próximos modelos, que são tidos como muito mais potentes, e que normalmente têm um desempenho mais poderoso.

### 3.1.4 Modelo dos Support Vector Machines

Novamente, assim como no caso do modelo autorregressivo e do KNN, mais de um modelo foi testado no caso dos SVM's. Por ser um modelo muito mais complexo do que o KNN, certamente seria mais difícil achar a melhor configuração neste caso. Nesse sentido, foram testados 3 diferentes modelos, com hiperparâmetros variados, diferentes números de vetores de suporte e kernels.

**Figura 9 – SVMs x IPCA**



O primeiro SVM foi feito simplesmente a partir do input dos dados para o software sem nenhuma tentativa de melhoria dos hiperparâmetros (o que não quer dizer que seja inferior, apenas que a própria função já avalia o melhor modelo automaticamente) e o resultado encontrado foi uma especificação com 70 vetores de suporte e kernel do tipo RBF (gaussiano). Já para o segundo SVM, foi realizada uma etapa de validação cruzada, permitindo um conjunto de valores para os principais hiperparâmetros, até encontrarmos a melhor combinação entre eles (que é uma combinação entre o parâmetro  $C$  – que penaliza os erros de classificação nos dados de teste – e um parâmetro  $\gamma$  – que diz respeito a quanto uma única observação interfere nos dados de treino). Assim, o que se obteve foi um modelo com 47 vetores de suporte, e um kernel do tipo linear. No terceiro SVM também foi feita validação cruzada, e o resultado foi um modelo com 65 vetores de suporte e kernel RBF.

A tabela 4 reúne todos os erros de previsão dos 3 SVM's para melhor visualização. Pela tabela, vê-se que, dentro desse conjunto de modelos, o SVM 1 é o que mostrou menor erro ao prever o IPCA, para todos os horizontes. Desta vez, ao realizar o teste de Diebold-Mariano,

identificou-se que os três modelos possuem acurácias iguais considerando todos os horizontes, com um p-valor muito próximo a 0.05 no período de 24 meses, mas ainda assim maior, de modo que eles não diferem. Apesar disso, para efeito de comparação com os demais modelos usados na dissertação, escolheu-se o SVM 1.

**Tabela 4 – Erros de previsão Support Vector Machines**

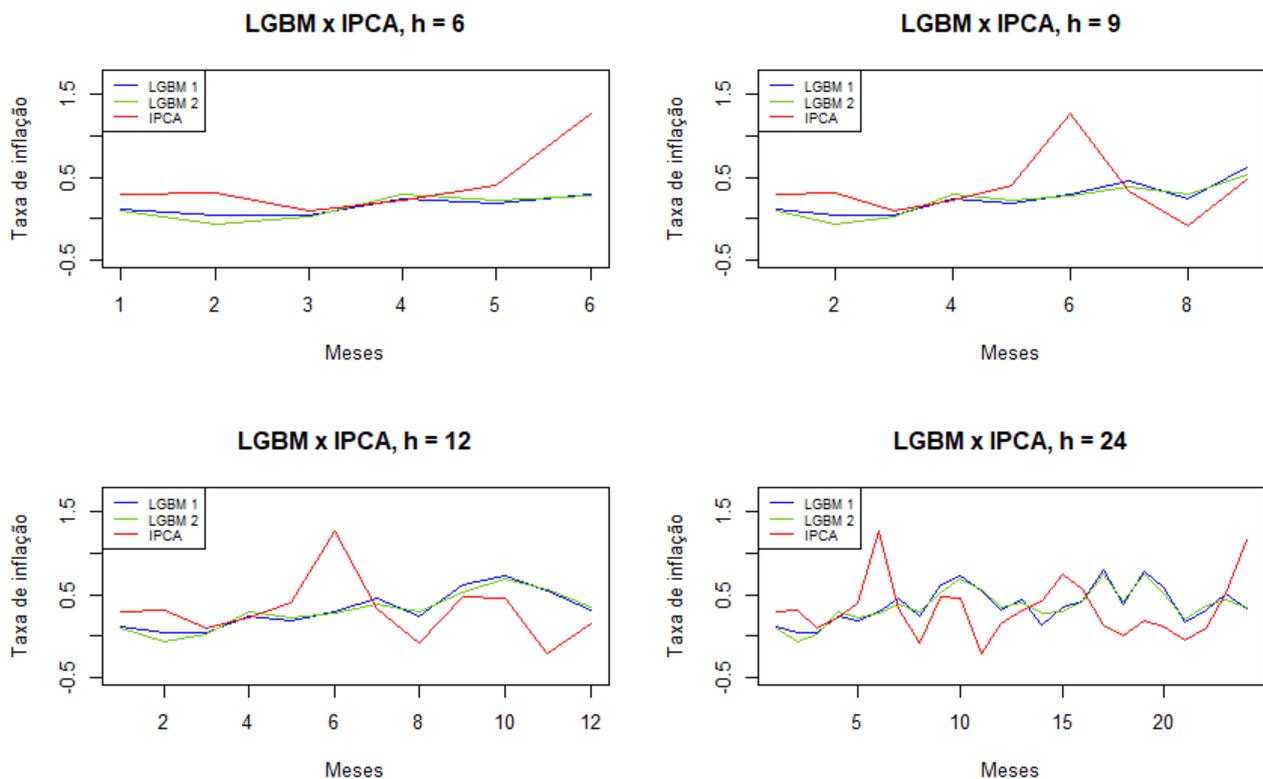
|              | Erros de previsão |           |           |           |           |           |           |           |
|--------------|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|              | 6 meses           |           | 9 meses   |           | 12 meses  |           | 24 meses  |           |
|              | MSE               | RMSE      | MSE       | RMSE      | MSE       | RMSE      | MSE       | RMSE      |
| <b>SVM 1</b> | 0.1602109         | 0.4002636 | 0.1363856 | 0.3693042 | 0.1537079 | 0.3920559 | 0.1599787 | 0.3999734 |
| <b>SVM 2</b> | 0.2449443         | 0.4949185 | 0.1842655 | 0.4292616 | 0.1981388 | 0.4451278 | 0.1938691 | 0.4403057 |
| <b>SVM 3</b> | 0.1682024         | 0.4101249 | 0.1386297 | 0.3723301 | 0.1576075 | 0.3969981 | 0.1653519 | 0.4066348 |

### 3.1.5 Modelo Light GBM

Finalmente, o último dos modelos utilizado na dissertação é oriundo da família das técnicas de gradient boosting, e bastante usado em competições de ciência de dados juntamente com o XGBoost. Como já explicado, o uso desse tipo de técnica aqui tem a ver com a grande capacidade de acurácia que ela costuma mostrar nessas competições em relação a outras técnicas. Assim, mais uma vez, foram testados mais de um modelo com diferentes valores nos hiperparâmetros. Com efeito, o primeiro modelo foi criado sem passar nenhum tipo de especificação para o método, apenas deixando o software realizar a implementação por si só, com os valores padrões/default. Já no segundo, alguns parâmetros foram alterados após uma etapa de validação cruzada, como o número de árvores (200; o default é 100). Em relação à taxa de aprendizado (*learning rate*),

que é um dos parâmetros mais importantes para algoritmos de *gradient boosting*, foi optado por deixar no default (0.1 – no geral, valores menores que 1 já são robustos contra *overfitting*)<sup>11</sup>. A taxa de aprendizado evita que o modelo gere *overfitting* ao aprender rápido demais a partir dos dados a cada nova árvore adicionada. Então, é recomendado que a taxa seja menor que 1, mas que geralmente costuma ficar dentro do intervalo de 0.1 a 0.3.

**Figura 10 – Light GBM x IPCA**



Tal como no caso dos SVM, graficamente, os modelos parecem ter um desempenho bastante semelhante entre si, sendo difícil identificar qual dos dois apresenta uma melhor performance. Para isso, a tabela 4 exibe os erros de previsão de cada um deles para os horizontes propostos.

<sup>11</sup> Apesar de ser uma explicação voltada ao XGBoost, no que tange ao *tuning* dos hiperparâmetros, há certa similaridade: <https://www.ibm.com/topics/xgboost>

Tabela 5 – Erros de previsão Light GBM

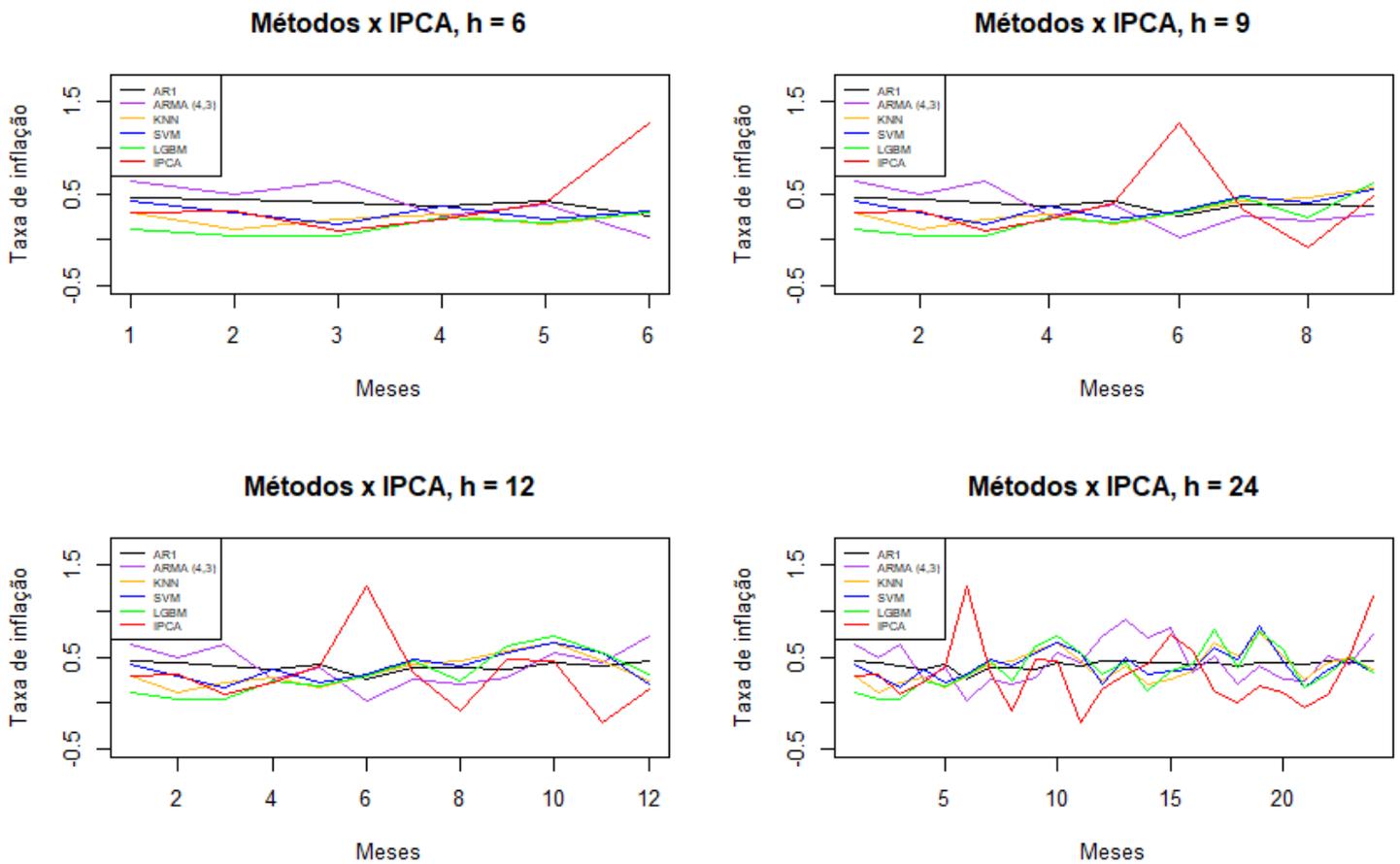
|               | Erros de previsão |           |           |           |           |           |           |           |
|---------------|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|               | 6 meses           |           | 9 meses   |           | 12 meses  |           | 24 meses  |           |
|               | MSE               | RMSE      | MSE       | RMSE      | MSE       | RMSE      | MSE       | RMSE      |
| <b>LGBM 1</b> | 0.1812349         | 0.4257169 | 0.1360511 | 0.368851  | 0.1576155 | 0.3970082 | 0.1708473 | 0.4133368 |
| <b>LGBM 2</b> | 0.1970167         | 0.4438657 | 0.1490875 | 0.3861185 | 0.1708555 | 0.4133467 | 0.1713317 | 0.4139223 |

Ao olhar a tabela, vê-se que, para quase todos os horizontes, o modelo Light GBM 1 apresenta um melhor desempenho. Novamente, antes de escolher um dos dois, é interessante usar o teste de Diebold-Mariano para saber se de fato essas diferenças expressam alguma significância estatística. Nesse sentido, mesmo com o tuning de hiperparâmetros, eles não apresentam diferença estatística em suas previsões. Porém, como um deles será escolhido para ser comparado com os demais modelos, optou-se pelo primeiro.

### 3.1.6 Comparação entre os principais modelos

A figura 11 traz todos os principais modelos escolhidos em uma comparação com o IPCA. Apesar da análise gráfica, a próxima seção reunirá os erros de previsão de cada uma delas, sendo um dos principais objetivos deste estudo.

Figura 11 – Métodos x IPCA



### 3.2 Erros de previsão

Nesta seção, será feita uma análise comparativa entre os desempenhos dos principais modelos de cada grupo. A tabela 6 abaixo traz essa comparação.

Em uma primeira análise em relação ao benchmark, os modelos de machine learning tiveram um resultado superior até 9 meses, tanto para o MSE, quanto para o RMSE. Para os horizontes de 12 e 24 meses, o desempenho destes modelos não foi exatamente o que se esperava deles, sobretudo por conta do prestígio que eles têm em previsões de longo prazo.

Tabela 6 – Erros de previsão dos modelos

|                   | Erros de previsão |           |           |           |           |           |           |           |
|-------------------|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|                   | 6 meses           |           | 9 meses   |           | 12 meses  |           | 24 meses  |           |
|                   | MSE               | RMSE      | MSE       | RMSE      | MSE       | RMSE      | MSE       | RMSE      |
| <b>AR 1</b>       | 0.1926843         | 0.4389582 | 0.1546484 | 0.3932536 | 0.1547431 | 0.393374  | 0.1349035 | 0.3672921 |
| <b>ARMA (4,3)</b> | 0.3300344         | 0.5744862 | 0.2356107 | 0.4853975 | 0.2394482 | 0.4893345 | 0.1676558 | 0.4094579 |
| <b>KNN</b>        | 0.1758133         | 0.419301  | 0.1522636 | 0.3902096 | 0.1559303 | 0.3948802 | 0.1684778 | 0.4104605 |
| <b>SVM 1</b>      | 0.1602109         | 0.4002636 | 0.1363856 | 0.3693042 | 0.1537079 | 0.3920559 | 0.1599787 | 0.3999734 |
| <b>LGBM 1</b>     | 0.1812349         | 0.4257169 | 0.1360511 | 0.368851  | 0.1576155 | 0.3970082 | 0.1708473 | 0.4133368 |

Dentre os modelos de séries de tempo (AR e ARMA), o AR (1) é o que apresenta os melhores resultados para todos os horizontes, com alguns dos erros menores até do que os modelos de aprendizado de máquina. Já para as técnicas de machine learning, os melhores modelos foram o SVM e o KNN. Entre os dois, o SVM teve uma performance melhor para todos os meses estipulados, embora essa diferença de desempenho só poderá ser balizada pelo teste de Diebold-Mariano na próxima seção. Mesmo assim, como o KNN é um algoritmo muito mais simples que os demais, foi surpreendente ele apresentar um desempenho tão parecido com os outros dois, que são mais rebuscados. É interessante ressaltar que o SVM é um dos modelos de machine learning mais complexos e poderosos que existem, mesmo que geralmente esse potencial seja mais explorado em problemas de classificação (Athey e Imbens, 2019). Também é importante dizer que o grande trunfo desse modelo acontece em bases de dados mais complicadas, em situações onde não é possível usar um kernel do tipo linear. Aqui, embora fosse possível usar um kernel linear conforme indicado pela validação cruzada (SVM 2), sem que se precisasse transformar o espaço das features, o modelo com kernel gaussiano (RBF) foi

escolhido, e desempenhou relativamente bem. Os resultados da técnica Light GBM também surpreenderam, só que negativamente. Dada a sua popularidade em competições de ciência de dados por conta do alto grau de acurácia que os algoritmos de Gradient Boosting obtêm, esperava-se que ele fosse um dos melhores, algo que não aconteceu para a base de dados aqui apresentada. Relembrando, porém, que esta dissertação tem como uma das inspirações o trabalho de Araújo & Gaglianone (2023), que usaram o XGBoost, que é o modelo mais famoso e usado dentre esse conjunto. Assim, para que não se repetisse o mesmo método, foi usado o Light GBM. De forma geral, nenhuma técnica usada naquele artigo foi repetida aqui, com exceção do benchmark. A título de ilustração, ao comparar o MSE relativo desta dissertação com o XGBoost do artigo mencionado, para os períodos de  $h = 6, 9$  e  $12$ , o que se obteve foi que o Light GBM gerou um MSE relativo de 39%, 22% e 23%, enquanto o XGBoost alcançou um MSE relativo de 21%, 20% e 18%.

### 3.3 Teste de Diebold-Mariano

A partir de agora, as próximas tabelas servirão para comparar e saber se os erros de previsão apresentados na seção anterior são de fato significativos em termos estatísticos comparando dois a dois. Os valores que estão nas tabelas são os p-valores para um nível de significância de 5%.

**Tabela 7 – Teste de Diebold-Mariano – AR x outros modelos**

|      | Teste Diebold-Mariano - AR x Modelos |            |         |        |        |
|------|--------------------------------------|------------|---------|--------|--------|
|      | h = 6 meses                          |            |         |        |        |
|      | AR 1                                 | ARMA (4,3) | KNN     | SVM 1  | LGBM 1 |
| AR 1 | -                                    | 0.1704     | 0.4365  | 0.2168 | 0.6683 |
|      | h = 9 meses                          |            |         |        |        |
| AR 1 | -                                    | 0.2395     | 0.8902  | 0.3045 | 0.3901 |
|      | h = 12 meses                         |            |         |        |        |
| AR 1 | -                                    | 0.1149     | 0.9426  | 0.9651 | 0.9079 |
|      | h = 24 meses                         |            |         |        |        |
| AR 1 | -                                    | 0.352      | 0.07522 | 0.2705 | 0.1654 |

Apesar de apresentar erros de previsão menores do que o ARMA (4,3) para todos os períodos analisados, e erros menores em relação a todos modelos de machine learning para o período e 24 meses (e 12 meses em comparação ao KNN e Light GBM), essa discrepância não se mostrou estatisticamente significativa, de modo que os erros do AR (1) são iguais a todos os outros quatro métodos.

**Tabela 8 – Teste de Diebold-Mariano – ARMA x outros**

|                   | <b>Teste Diebold-Mariano - ARMA x Modelos</b> |                   |            |              |               |
|-------------------|---|-------------------|------------|--------------|---------------|
|                   | <b>h = 6 meses</b>                            |                   |            |              |               |
|                   | <b>AR 1</b>                                   | <b>ARMA (4,3)</b> | <b>KNN</b> | <b>SVM 1</b> | <b>LGBM 1</b> |
| <b>ARMA (4,3)</b> | 0.1704  | -                 | 0.2341     | 0.173        | 0.2117        |
|                   | <b>h = 9 meses</b>                            |                   |            |              |               |
| <b>ARMA (4,3)</b> | 0.2395  | -                 | 0.1845     | 0.2452       | 0.2016        |
|                   | <b>h = 12 meses</b>                           |                   |            |              |               |
| <b>ARMA (4,3)</b> | 0.1149  | -                 | 0.07313    | 0.2223       | 0.2073        |
|                   | <b>h = 24 meses</b>                           |                   |            |              |               |
| <b>ARMA (4,3)</b> | 0.352   | -                 | 0.5538     | 0.8732       | 0.9461        |

Em relação ao modelo ARMA (4,3), ainda que ele tenha sido o pior modelo de todos no que se refere aos erros de previsão, os resultados do teste DM indicam que, mesmo assim, ele não se diferencia dos demais por conta disso, em termos estatísticos.

**Tabela 9 – Teste de Diebold-Mariano – KNN x outros modelos**

|            | <b>Teste Diebold-Mariano - KNN x Modelos</b> |                   |            |              |               |
|------------|--|-------------------|------------|--------------|---------------|
|            | <b>h = 6 meses</b>                           |                   |            |              |               |
|            | <b>AR 1</b>                                  | <b>ARMA (4,3)</b> | <b>KNN</b> | <b>SVM 1</b> | <b>LGBM 1</b> |
| <b>KNN</b> | 0.4365                                       | 0.2341            | -          | 0.2542       | 0.5944        |
|            | <b>h = 9 meses</b>                           |                   |            |              |               |
| <b>KNN</b> | 0.8902                                       | 0.1845            | -          | 0.1567       | 0.5068        |
|            | <b>h = 12 meses</b>                          |                   |            |              |               |
| <b>KNN</b> | 0.9426                                       | 0.07313           | -          | 0.8713       | 0.9358        |
|            | <b>h = 24 meses</b>                          |                   |            |              |               |
| <b>KNN</b> | 0.07522                                      | 0.5538            | -          | 0.3932       | 0.8729        |

Para o KNN, embora ele tenha apresentado um p-valor baixo (0.07) se comparado ao ARMA (4,3) para 12 meses, e o AR (1) para 24 meses, também não foi suficiente para que ele apresentasse alguma diferença estatística em relação a esses modelos, e nem qualquer outro.

**Tabela 10 – Teste de Diebold-Mariano – SVM x outros modelos**

|              | <b>Teste Diebold-Mariano - SVM 2 x Modelos</b> |                   |            |              |               |
|--------------|--|-------------------|------------|--------------|---------------|
|              | <b>h = 6 meses</b>                             |                   |            |              |               |
|              | <b>AR 1</b>                                    | <b>ARMA (4,3)</b> | <b>KNN</b> | <b>SVM 1</b> | <b>LGBM 1</b> |
| <b>SVM 1</b> | 0.2168   | 0.173             | 0.2542     | -            | 0.2103        |
|              | <b>h = 9 meses</b>                             |                   |            |              |               |
| <b>SVM 1</b> | 0.3045   | 0.2452            | 0.1567     | -            | 0.9868        |
|              | <b>h = 12 meses</b>                            |                   |            |              |               |
| <b>SVM 1</b> | 0.9651   | 0.2223            | 0.8713     | -            | 0.7988        |
|              | <b>h = 24 meses</b>                            |                   |            |              |               |
| <b>SVM 1</b> | 0.2705   | 0.8732            | 0.3932     | -            | 0.4391        |

Já para o SVM, ainda que seja o modelo de machine learning com o melhor resultado, além de superar o AR (1) em quase todos os períodos, o teste DM também apontou para uma não significância estatística em relação aos demais modelos.

**Tabela 11 – Teste de Diebold-Mariano – Light GBM x outros modelos**

|               | <b>Teste Diebold-Mariano - LGBM 2 x Modelos</b> |                   |            |              |               |
|---------------|---|-------------------|------------|--------------|---------------|
|               | <b>h = 6 meses</b>                              |                   |            |              |               |
|               | <b>AR 1</b>                                     | <b>ARMA (4,3)</b> | <b>KNN</b> | <b>SVM 1</b> | <b>LGBM 1</b> |
| <b>LGBM 1</b> | 0.6683  | 0.2117            | 0.5944     | 0.2103       | -             |
|               | <b>h = 9 meses</b>                              |                   |            |              |               |
| <b>LGBM 1</b> | 0.3901  | 0.2016            | 0.5068     | 0.9868       | -             |
|               | <b>h = 12 meses</b>                             |                   |            |              |               |
| <b>LGBM 1</b> | 0.9079  | 0.2073            | 0.9358     | 0.7988       | -             |
|               | <b>h = 24 meses</b>                             |                   |            |              |               |
| <b>LGBM 1</b> | 0.1654  | 0.9461            | 0.8729     | 0.4391       | -             |

Pela tabela 11, o Light GBM também não apresentou diferença estatística em relação aos outros modelos.

Dessa maneira, ao olhar os p-valores em cada uma das tabelas, algumas conclusões podem ser tiradas. A primeira delas, no que diz respeito ao benchmark, é que, apesar de ele ter tido um desempenho melhor do que o ARMA, e semelhante (e até superior) aos de machine learning para alguns dos períodos, essa divergência no MSE e RMSE não se ratificou ao fazer o teste de Diebold-Mariano.

Em relação ao modelo ARMA (4,3), mesmo sendo o pior entre todos os modelos, o teste DM mostrou que essa diferença entre os erros de previsão não foi suficiente para que ela se validasse em termos estatísticos.

Agora, para os modelos de machine learning, ainda que o KNN seja um modelo inferior aos outros, os erros de previsão gerados por ele não apresentaram diferença estatística em relação ao SVM e ao Light GBM para qualquer horizonte, isto é, mesmo que ele tenha surpreendido ao desempenhar melhor do que o Light GBM, e mesmo que tenha sido inferior ao SVM para qualquer horizonte, essas diferenças não se mostraram relevantes. Finalmente, comparando as duas técnicas mais sofisticadas, o que aconteceu mais uma vez é que os erros de previsão entre elas não exibiram diferença estatística, ainda que esse resultado em particular talvez fizesse algum sentido na medida em que havia uma expectativa de que elas viriam a ter um resultado melhor do que os demais modelos, e que, portanto, talvez seus desempenhos fossem mais ou menos parecidos, especialmente no longo prazo.

Então, podemos concluir que, em relação ao erro de previsão puramente, considerando os modelos de machine learning, tivemos os seguintes resultados: ao se considerar apenas o KNN e o Light GBM, eles apresentaram uma acurácia maior para 6 e 9 meses em relação ao benchmark, enquanto o AR (1) teve um desempenho melhor para 12 e 24 meses. Já o SVM, também foi superior em 12 meses, mas, no período de 24 meses, também foi inferior ao AR. Entretanto, ao realizar o teste de Diebold-Mariano para confirmar esses resultados, percebeu-se que eles eram iguais em termos estatísticos.

## CONCLUSÃO

A importância dos agentes econômicos em ter algum tipo de previsibilidade sobre o nível de preços reforça a necessidade de exercícios de previsão sobre a variação da taxa de inflação de um determinado país. Para realizar essa difícil tarefa, são usados desde modelos mais simples oriundos da econometria, até modelos mais rebuscados de diferentes especificações e origens, frequentemente usados por instituições financeiras e o próprio Banco Central.

Nesta dissertação, tentou-se, por dois grupos de técnicas, prever a variação da inflação brasileira representada pelo índice de preços ao consumidor amplo (IPCA). Foram empregados tanto métodos mais tradicionais, assim como outros que vêm crescendo em popularidade nos últimos anos, que são os métodos de *machine learning*.

Olhando meramente para os desempenhos individuais de cada um dos modelos em termos de erros de previsão, verificou-se que uma técnica computacional altamente robusta, o Support Vector Machine, teve um resultado superior à maioria dos modelos para todos os períodos de previsão, com exceção do AR (1). Ao compará-lo com os modelos de séries temporais mais clássicos, a diferença realmente é notória em relação ao ARMA (4,3), embora não tenha conseguido performar melhor do que o AR (1) no horizonte de 24 meses. Já no contraste com as duas outras técnicas de aprendizado de máquina, ainda que ela tenha se saído melhor para qualquer período, essa diferença não é tão nítida assim, no sentido de que os erros são muito próximos entre si. Como se sabe, o objetivo dos modelos de machine learning é a previsão, tanto para problemas de classificação e de regressão.

De certo modo, o uso dessas técnicas mais modernas trazia uma grande expectativa de bons resultados, principalmente no que concerne o SVM e o Light GBM. Todavia, notou-se que o modelo ingênuo (AR 1) foi superior a todos os modelos de machine learning para o período de 24 meses, superando também o KNN e o Light GBM em 12 meses. De certa forma, esse desfecho não é uma novidade para a série de inflação brasileira, uma vez que esse também foi o resultado obtido por Medeiros, Vasconcelos e Freitas (2016). Ou seja, mesmo usando uma base de dados com dezenas de variáveis, e muitos métodos de previsão com um grau de sofisticação mais alto, o melhor modelo para longos horizontes foi o autorregressivo. Isso ocorre pelo fato de que a inflação, no longo prazo, só é afetada verdadeiramente por ela própria

em virtude da inércia, e por variáveis relacionadas à dívida pública, de forma que o resto das variáveis são pouco relevantes.

Posteriormente, ao realizar o teste estatístico de Diebold-Mariano para cancelar as eventuais superioridades de previsão, concluiu-se que não houve diferença estatística entre os métodos empregados neste trabalho,

Em um sentido mais amplo, a tendência é que os algoritmos oriundos do aprendizado de máquina sejam cada vez mais usados para prever variáveis macroeconômicas, além de uma combinação entre esses modelos, como fizeram Araújo & Gaglianone (2023), já que uma agregação de técnicas costuma performar melhor em relação a um único modelo. Quando se leva em conta o kit de ferramentas usadas por agentes de instituições financeiras e do Banco Central, é um outro universo, tanto pela quantidade de recursos investida para essa tarefa, quanto pelo maior conjunto de informações que eles possuem, especialmente o Banco Central. Assim, a junção desses recursos e informações com modelos bem fundamentados aumenta substancialmente a qualidade da análise gerada por essas corporações. Entretanto, também é inegável que a tentativa de prever uma variável como a inflação sempre será muito difícil por conta da incerteza existente na economia, independentemente do material usado.

Por fim, esta dissertação buscou, de forma modesta, contribuir em alguma dimensão para essa importante atividade econômica, tomando como inspiração trabalhos anteriores nacionais e internacionais, e fica como uma possível contribuição para obras futuras.

## REFERÊNCIAS BIBLIOGRÁFICAS

ADRIAN, Tobias. The Role of Inflation Expectations in Monetary Policy. **IMF Financial Counsellor and Director of the Monetary and Capital Markets Department**, 2023.

Disponível em: <https://www.imf.org/en/News/Articles/2023/05/15/sp-role-inflation-expectations-monetary-policy-tobias-adrian> Acesso em: 22 de maio de 2024.

ARAÚJO, Gustavo Silva; GAGLIANONE, Wagner Piazza. Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, v. 4, n. 2, p. 100087, 2023.

ARRUDA, E. F., FERREIRA, R. T., & CASTELAR, I. (2011). Modelos lineares e não lineares da curva de Phillips para previsão da taxa de inflação no Brasil. *Revista Brasileira de Economia*, 65, 237-252.

ATHEY, Susan; IMBENS, Guido W. Machine learning methods that economists should know about. **Annual Review of Economics**, v. 11, n. 1, p. 685-725, 2019.

ATKESON, Andrew et al. Are Phillips curves useful for forecasting inflation?. **Federal Reserve bank of Minneapolis quarterly review**, v. 25, n. 1, p. 2-11, 2001.

Banco Central do Brasil (BCB). Núcleo de inflação, junho 2000. Disponível em: <ri200006b4p.pdf> ([bcb.gov.br](http://www.bcb.gov.br)) Acesso em: 16 de abril de 2024.

Banco Central do Brasil (BCB). Relatório de inflação, Volume 12, número 1, 1-157, março de 2010. Disponível em: <https://www.bcb.gov.br/htms/relinf/port/2010/03/ri201003inp.pdf>. Acesso em: 11 de abril de 2024.

Banco Central do Brasil (BCB). Novas medidas de núcleo de inflação, junho 2018. Disponível em:

[https://www.bcb.gov.br/conteudo/relatorioinflacao/estudosoespeciais/novas\\_medidas\\_nucleo\\_inflacao.pdf](https://www.bcb.gov.br/conteudo/relatorioinflacao/estudosoespeciais/novas_medidas_nucleo_inflacao.pdf) Acesso em: 19 de abril de 2024.

Banco Central do Brasil (BCB). Modelo de vetor autorregressivo com ancoragem de longo prazo, junho 2018. Disponível em: [Modelo de vetor autorregressivo com ancoragem de longo prazo](#) Acesso em: 14 de setembro de 2024.

Banco Central do Brasil (BCB). Relatório de inflação, Volume 26, número 2, 1-114, junho de 2024. Disponível em: <https://www.bcb.gov.br/publicacoes/ri/202406> Acesso em: 09 de julho de 2024.

BLANCHARD, Olivier. *Macroeconomia*. 5ª ed. São Paulo: Prentice Hall, 2010.

BUENO, Rodrigo de Losso da Silveira. *Econometria de séries temporais*. 2018.

CASTELAR, Ivan; SANTOS, Cristiano. Avaliando e Propondo Medidas de Núcleo da Inflação no Brasil. **Série Estudos Econômicos CAEN**, n. 10, 2016.

CHAKRABORTY, Chiranjit; JOSEPH, Andreas. *Machine learning at central banks*. 2017.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost. A Scalable Tree Boosting system. In: **Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**. 2016. p. 785-94.

COIBION, Olivier; GORODNICHENKO, Yuriy; ROPELE, Tiziano. Inflation expectations and firm decisions: New causal evidence. **The Quarterly Journal of Economics**, v. 135, n. 1, p. 165-219, 2020.

CORTES, Corinna; VAPNIK, Vladimir. Support-Vector Networks. **Machine Learning**, 1995.

DA SILVA FILHO, Tito Nícias Teixeira et al. Revisitando as Medidas de Núcleo de Inflação do Banco Central do Brasil. **Banco Central do Brasil-Trabalhos para Discussão**, v. 356, 2014.

DIEBOLD, Francis X.; MARIANO, Robert S. Comparing predictive accuracy. **Journal of Business & economic statistics**, v. 20, n. 1, p. 134-144, 2002.

ELLIOTT, Graham; GRANGER, Clive; TIMMERMANN, Allan (Ed.). **Handbook of economic forecasting. Volume 1**. Elsevier, 2006..

ELLIOTT, Graham; TIMMERMANN, Allan (Ed.). **Handbook of economic forecasting. Volume 2A**. Elsevier, 2013.

EL-SHAGI, Makram; GIESEN, Sebastian; JUNG, Alexander. Does the federal reserve staff still beat private forecasters?. 2014.

EINAV, Liran; LEVIN, Jonathan. Economics in the age of big data. **Science**, v. 346, n. 6210, p. 1243089, 2014.

JAMES, Gareth et. al. An introduction to statistical learning. 2013.

JOUILIL, Youness et al. Comparing the accuracy of classical and machine learning methods in time series forecasting: A case study of usa inflation. **Statistics, Optimization & Information Computing**, v. 11, n. 4, p. 1041-1050, 2023.

IVAŞCU, Codruţ. Can Machine Learning Models Predict Inflation?. In: **Proceedings of the International Conference on Business Excellence**. 2023. p. 1748-1756.

KAIHATSU, Sohei; SHIRAKI, Noriyuki. **Firms' inflation expectations and wage-setting behaviors**. Bank of Japan, 2016.

KE, Guolin et al. LightGBM: A highly efficient gradient boosting decision tree. **Advances in neural information processing systems**, v. 30, 2017.

LACK, Caesar. **Forecasting Swiss inflation using VAR models**. Swiss National Bank, 2006.

RESENDE, André Lara. **Juros, moeda e ortodoxia: teorias monetárias e controvérsias políticas**. Portfolio-Penguin, 2017.

LAVOIE, Marc. Post-Keynesian economics: new foundations. In: **Post-Keynesian Economics**. Edward Elgar Publishing, 2014.

LIPOVINA-BOŽOVIĆ, Milena; JU, Cerović; VUJOŠEVIĆ, Saša. Forecasting inflation in Montenegro using univariate time series models. **Business and Economic Horizons (BEH)**, v. 11, n. 1, p. 51-63, 2015.

MASON, Llew et al. Boosting algorithms as gradient descent. **Advances in neural information processing systems**, v. 12, 1999.

MARTIN, Vance; HURN, Stan; HARRIS, David. **Econometric modelling with time series: specification, estimation and testing**. Cambridge University Press, 2013.

MEDEIROS, Marcelo C.; VASCONCELOS, Gabriel; FREITAS, Eduardo. Forecasting brazilian inflation with high-dimensional models. **Brazilian Review of Econometrics**, v. 36, n. 2, p. 223-254, 2016

MENGUS, Eric; ANDRADE, Philippe; GAUTIER, Erwan. **What Matters in Households' Inflation Expectations?**. CEPR Discussion Papers, 2020.

MISHKIN, Frederic S. Inflation dynamics. **International Finance**, v. 10, n. 3, p. 317-334, 2007.

RASCHKA, Sebastian. STAT 479: Machine Learning Lecture Notes. Department of Statistics, University of Wisconsin-Madison. Disponível em: [https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02\\_knn\\_notes.pdf](https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02_knn_notes.pdf) Acesso em: 16 de setembro de 2024.

ROBINSON, Joan. **Essays in the theory of economic growth**. Springer, 1962.

RODRÍGUEZ-VARGAS, Adolfo. Forecasting Costa Rican inflation with machine learning methods. **Latin American Journal of Central Banking**, v. 1, n. 1-4, p. 100012, 2020.

ROGER, Scott. Core inflation: concepts, uses and measurement. **Reserve Bank of New Zealand Discussion Paper**, n. G98/9, 1998.

RUCH, Franz et al. Forecasting core inflation: The case of South Africa. **Applied Economics**, v. 52, n. 28, p. 3004-3022, 2020.

STOCK, James H.; WATSON, Mark W. Forecasting inflation. **Journal of monetary economics**, v. 44, n. 2, p. 293-335, 1999.

STOCK, James H.; WATSON, Mark W. Why has US inflation become harder to forecast?. **Journal of Money, Credit and banking**, v. 39, p. 3-33, 2007.

SVENSSON, Lars EO. Inflation forecast targeting: Implementing and monitoring inflation targets. **European economic review**, v. 41, n. 6, p. 1111-1146, 1997.

SVENSSON, Lars EO. Monetary policy with judgment: Forecast targeting. 2005.

TULIP, Peter. Has the economy become more predictable? Changes in Greenbook forecast accuracy. **Journal of Money, Credit and Banking**, v. 41, n. 6, p. 1217-1231, 2009.

VARIAN, Hal R. Big data: New tricks for econometrics. **Journal of economic perspectives**, v. 28, n. 2, p. 3-28, 2014.

WEINTRAUB, Sidney. Capitalism's inflation and unemployment crisis: beyond monetarism and Keynesianism, 1978.

WRAY, L. Modern Money Theory: A Primer on Macroeconomics for Sovereign Monetary Systems (2ª Edição ed.). **Nova Iorque: Palgrave**, 2015.

WYNNE, Mark A. Core inflation: a review of some conceptual issues. Available at SSRN 355143, 1999.