



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO SÓCIOECONÔMICO  
CURSO DE GRADUAÇÃO EM CIÊNCIAS ECONÔMICAS

Wellington Gubiani Maiberg

**Previsão em tempo real da Inflação Brasileira com aprendizado de máquina**

Florianópolis/SC

2022

PREVISÃO EM TEMPO REAL DA INFLAÇÃO BRASILEIRA COM APRENDIZADO DE  
MÁQUINA

Wellington Gubiani Maiberg

Trabalho de Conclusão de Curso da Graduação  
em Ciências Econômicas do Centro Sócioeconômico da Universidade Federal de Santa Catarina  
para a obtenção do título de bacharel em Ciências Econômicas.

Orientador: Guilherme Valle Moura

Florianópolis/SC

2022

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

MAIBERG, WELLINGTON

Previsão em tempo real da Inflação Brasileira com  
aprendizado de máquina / WELLINGTON MAIBERG ; orientador,  
Guilherme Valle Moura, 2022.

24 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Centro Sócio  
Econômico, Graduação em Ciências Econômicas, Florianópolis,  
2022.

Inclui referências.

1. Ciências Econômicas. 2. Nowcasting. 3. Inflação. 4.  
Aprendizado de Máquina. 5. Previsão. I. Valle Moura,  
Guilherme . II. Universidade Federal de Santa Catarina.  
Graduação em Ciências Econômicas. III. Título.

**Wellington Gubiani Maiberg**

**Previsão em tempo real da Inflação Brasileira com aprendizado de máquina**

O presente trabalho em nível de graduação foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Guilherme Valle Moura, Dr.  
Universidade Federal de Santa Catarina

Prof. Roberto Meurer, Dr.  
Universidade Federal de Santa Catarina

Prof. Pedro Luiz Paolino Chaim, Dr.  
Universidade Federal de Santa Catarina

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Ciências Econômicas” e aprovado em sua forma final pelo Curso de Graduação em Ciências Econômicas.

---

Prof.(a) Guilherme Valle Moura, Dr.(a)  
Orientador

Florianópolis, 2022.

Este trabalho é dedicado aos meus colegas de universidade e aos meus queridos familiares que tanto acreditaram em mim.

# Agradecimentos

Gostaria de agradecer aos colegas Richard Schnorrenberger e Aishameriane Venes Schmidt pelo conhecimento e paciência ao me guiarem neste processo de aprendizado, a todos os amigos que de alguma forma me ajudaram, seja com incentivo, seja com colaborações, e principalmente agradecer ao meu orientador, Guilherme Valle Moura, que além de me apresentar ao tema, sempre conseguia me inspirar com a sagacidade que explicava algo.

# Resumo

O presente trabalho tem como proposta a realização de previsões em tempo real (*Nowcasting*) da inflação brasileira. A inflação quando baixa, estável e previsível pode trazer vários benefícios para a sociedade. A incerteza na economia diminui e os indivíduos podem planejar seu futuro sem ver sua renda sendo corroída. Utilizaremos métodos de regressão com aprendizado de máquina que realizam seleção e regularização de variáveis para melhorar a precisão da previsão e a interpretabilidade do modelo. Para que isso seja possível, é necessário a criação de uma base de dados de alta frequência. A fim de observar a acurácia, aplica-se uma avaliação de capacidade preditiva em tempo real.

**Palavras-chave:** Inflação. Nowcasting da inflação. Aprendizado de máquina.

# Abstract

The present work proposes the realization of forecasts in real time (Nowcasting) of Brazilian inflation. Inflation when low, stable and predictable can bring several benefits to society. Uncertainty in the economy decreases and individuals can plan their future without seeing their income erode. We will use regression methods with machine learning that perform selection and regularization of variables to improve prediction accuracy and model interpretability. For this to be possible, it is necessary to create a high frequency database. In order to observe the accuracy, A real-time predictive capability assessment is applied

**Keywords:** Inflation. Inflation Nowcasting. Machine Learning.

# Lista de ilustrações

Figura 1 – Dados mensais preço . . . . .	8
Figura 2 – Dados mensais atividade econômica . . . . .	9
Figura 3 – Dados semanais combustíveis . . . . .	9
Figura 4 – Dados semanais espec . . . . .	9
Figura 5 – Dados de frequência diária . . . . .	10
Figura 6 – Gráfico de calor RIDGE . . . . .	19
Figura 7 – Gráfico de calor LASSO . . . . .	20
Figura 8 – Gráfico de calor Elastic Net . . . . .	21

# Lista de quadros

# Lista de tabelas

Tabela 1 – RMSE solução AR para Raged Edge . . . . .	18
Tabela 2 – RMSE solução RW para Raged Edge . . . . .	22

# Lista de abreviaturas e siglas

PIB : Produto Interno Bruto

NAIRU : non-accelerating inflation rate of unemployment

IPCA : Índice Nacional de Preços ao Consumidor Amplo

IBGE : Instituto Brasileiro de Geografia e Estatística

LASSO: Least Absolute Shrinkage and Selection Operator

SRQ : Soma dos resíduos ao quadrado

AR1 : Auto Regressivo de primeira ordem

IPCS : Índice de Preços ao Consumidor

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>2</b>	<b>DESENVOLVIMENTO</b>	<b>2</b>
2.1	Por que prever a inflação?	2
2.2	Introdução ao nowcasting	3
2.3	Dilema Viés e Variância	4
<b>3</b>	<b>DADOS</b>	<b>7</b>
3.1	Janela Móvel	8
3.2	Apresentação dos regressores escolhidos	8
<b>4</b>	<b>METODOLOGIA</b>	<b>11</b>
<b>4.1</b>	<b>Regressão Linear</b>	<b>11</b>
4.1.1	Multicolinearidade e Sobreajuste	11
<b>4.2</b>	<b>Regressão RIDGE, LASSO e Elastic Net</b>	<b>12</b>
4.2.1	Introdução à regularização	12
4.2.2	Regressão Ridge	12
4.2.3	LASSO	13
4.2.4	Elastic Net	13
<b>4.3</b>	<b>Avaliação de performance do modelo</b>	<b>13</b>
<b>4.4</b>	<b>Modelo de frequência mista irrestrito</b>	<b>14</b>
4.4.1	Modelo de frequência mista irrestrito em notação matricial	15
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>16</b>
<b>5.1</b>	<b>Comparando a performance dos modelos</b>	<b>16</b>
5.1.1	Solução Autoregressiva para <i>ragged edge</i>	16
5.1.1.1	Método dos mínimos quadrados	16
5.1.1.2	Método de regularização RIDGE	16
5.1.1.3	Método de regularização LASSO	17
5.1.1.4	Método de regularização Elastic Net	17
5.1.1.5	Expectativas do mercado Boletim Focus	17
5.1.2	Solução de passeio aleatório para <i>ragged edge</i>	17
5.1.2.1	Método dos mínimos quadrados	18
5.1.2.2	Método de regularização RIDGE	18
5.1.2.3	Método de regularização LASSO	19
5.1.2.4	Método de regularização Elastic Net	20

5.1.2.5	Expectativas Boletim Focus . . . . .	21
<b>6</b>	<b>CONCLUSÃO</b> . . . . .	<b>23</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>24</b>

# 1 Introdução

Este trabalho propõe a realização de previsão em tempo real da inflação brasileira. Esse tipo de método tem por característica utilizar-se de uma gama de dados que são divulgados com diferentes datas e frequências. As vantagens em se utilizar um modelo que possa agregar esse tipo de dados se dá a partir da limitação gerada por séries de baixa frequência, como o PIB que é divulgado trimestralmente, e neste trabalho, a inflação, que é divulgada mensalmente. A maior dificuldade para o caso brasileiro é construir bases de dados, tendo em vista que estes são divulgados por diferentes instituições e possuem diferentes datas de referência e publicação. Uma ferramenta que incorpora as informações de forma mais atualizada, em um ambiente em que os dados são liberados de forma não síncrona, e com diferentes defasagens, é chamada *nowcasting*. Para realizar o *nowcasting*, precisamos apontar quais indicadores econômicos apresentam maior relevância para a previsão da inflação e organizá-los em uma base de dados com suas datas de divulgação e referência.

Para a realização do *nowcasting*, serão utilizados o modelo clássico de regressão linear e modelos com regularização tradicionais na literatura de aprendizado de máquinas. Para podermos inferir se há utilidade em aplicarmos esses métodos na previsão da inflação, as análises empíricas são comparadas às previsões com passeio aleatório e média das expectativas de mercado do Relatório Focus para o período de 2013 a 2021.

A relevância do tema proposto se dá a partir importância em se obter um modelo capaz de capturar com maior relevância os choques inflacionários e a necessidade em monitorar o comportamento da inflação, tanto para o setor público que pode responder com políticas monetárias, quanto para o setor privado que, por sua vez, necessita constantemente atualizar seus preços além de influenciar nas decisões dos agentes como: consumo, investimento e poupança. A inflação pode decorrer de um desequilíbrio entre oferta e demanda, uma estimativa em tempo real pode auxiliar o mercado a entrar em equilíbrio, além de controlar as expectativas.

O presente estudo está dividido em seis etapas, sendo elas: Introdução, na qual é ressaltada brevemente a importância do *Nowcasting*, segunda parte sendo desenvolvimento, abordando citações, referências e relato de experiências em estudos que alicerçam referenciais teóricos e revisão da literatura, terceira parte sobre a base de dados e suas transformações, quarta parte fala sobre a metodologia aplicada e os modelos escolhidos, quinta parte onde são apresentados os resultados e por último a conclusão, colocando considerações finais sobre o tema aqui presente.

## 2 Desenvolvimento

### 2.1 Por que prever a inflação?

Prever a inflação é uma preocupação tanto para profissionais do mercado quanto para bancos centrais. Os bancos centrais tem como responsabilidade garantir a estabilidade econômica do país, por meio da manutenção do poder de compra da moeda e da regulação do sistema financeiro, sendo assim, necessitam monitorar constantemente a inflação. Já os agentes do mercado necessitam atualizar constantemente suas expectativas, e ao passo que novas informações são divulgadas surgem oportunidades ou “riscos”, sendo assim também possuem um forte interesse em monitorar a inflação.

A partir da descoberta de William Phillips, a curva de Phillips, que representa uma relação inversa previsível entre inflação e desemprego no curto prazo, se o desemprego caísse, a inflação aumentaria e vice-versa. Economistas vêm usando-a como ferramenta principal para previsão da Inflação desde então, e aprimorando seus modelos baseados nela. A maioria gira em torno de uma taxa de desemprego ideal, a famosa NAIRU ou Taxa de Desemprego Não Aceleradora da Inflação. Na década de 70 Milton Friedman colocava em xeque seus conceitos teóricos e posteriormente, Stock e Watson (2008) trouxeram evidências consideráveis de que a curva de Phillips não traz melhorias para as previsões frente aos modelos de passeio aleatório quando a taxa de desemprego se distancia da NAIRU. Isso nos traz de volta ao problema inicial, no qual precisamos de um modelo melhor e mais atualizado para prever a inflação frente a cenários de inflação alta, como no Brasil, que apresenta no presente, alta taxa de desemprego e inflação.

O Brasil sofreu fortemente com a inflação pré-real, que pôde ser controlada somente quando o país passou a oferecer a paridade dólar-real, como estratégia de âncora cambial de política monetária, quando da implementação do Plano Real. Essa medida levou o país a um certo controle inflacionário, mas como apontam economistas, o preço a se pagar pode ser relativamente alto no controle da demanda por dólar. Quando essa política se torna obsoleta, e o país deixa o câmbio flutuante, a necessidade por outra âncora surge, e assim inicia o regime de metas da inflação.

A principal métrica da inflação brasileira é o Índice Nacional de Preços ao Consumidor Amplo (IPCA). O *nowcasting* é particularmente relevante para o IPCA, que é divulgado mensalmente, e lançado com atraso de um mês e alguns dias. O IPCA mede a variação de preços de uma cesta de produtos e serviços consumida pela população. O resultado mostra se os preços aumentaram ou diminuíram de um mês para o outro. A cesta é definida pela Pesquisa de Orçamentos Familiares do IBGE, a qual, dentre outras

questões, verifica o que a população consome e quanto do rendimento familiar é gasto em cada produto: arroz, feijão, passagem de ônibus, material escolar, médico, cinema, entre outros. O índice, portanto, leva em conta não apenas a variação de preço de cada item, mas também o peso que ele tem no orçamento das famílias.

## 2.2 Introdução ao nowcasting

O termo *nowcasting* já vem sendo usado há um tempo em meteorologia, e mais recentemente foi adotado pelos Economistas. O termo se refere a *now* como uma definição do tempo recente (agora), e *forecasting*, que seria previsão. Sendo assim, nowcasting traz a ideia de uma previsão para o passado próximo, o tempo recente e o futuro no curto prazo.

O princípio básico do *nowcasting* é a exploração da informação que é publicada antecipadamente e possivelmente em frequências mais altas do que a variável alvo de interesse, a fim de obter uma “estimativa antecipada” antes que o valor oficial esteja disponível (BAÑBURA et al., 2013). Para o caso aqui exposto, usaremos variáveis que são liberados com antecedência ao IPCA, com divulgações independentes e assimétricas.

Utilizar dados de alta frequência traz ganhos de previsão, além de maior flexibilidade do modelo ao se adaptar com a atualização de previsões devido ao fluxo constante de dados sendo incluídos ao modelo quando de suas divulgações. Isso pode ser observado no trabalho de Ghysels et al. (2012), no qual é utilizado uma série que envolve o conjunto típico de *commodities*, Série Risco Corporativo, Ações, Câmbio e Títulos Governamentais para realizar o *nowcasting* do crescimento do PIB, que performou melhor quando comparado a modelos tradicionais de dados agregados.

Com os dados entrando de forma não síncrona, utiliza-se um filtro de Kalman, que gera novas previsões para todas as variáveis consideradas e permite calcular para cada data de divulgação um modelo acolhendo as “novidades”. A revisão agora lançada pode ser considerada como uma média ponderada dessas novidades. Dessa forma os lançamentos de dados podem ser modelados e os diferentes indicadores podem nos trazer informações a respeito da atividade econômica. Este tipo de modelo é mencionado em Giannone, Reichlin e Small (2008) dentre outros posteriores.

Uma alternativa aos modelos usando o Filtro de Kalman é apresentada em Babii, Ghysels e Striaukas (2021), no qual é utilizado o modelo LASSO para séries temporais de frequência mista. Como a Regressão Lasso pode reduzir os parâmetros até zero, fazendo com que esses desapareçam, é mais indicado do que a Regressão Ridge para reduzir a variância em modelos que contêm muitas variáveis correlacionadas.

Os primeiros trabalhos de *nowcasting* foram direcionados ao PIB e posteriormente para alvos de maior frequência. Para executar o *nowcasting*, buscam-se informações que

apresentem correlação com o alvo e tenham uma frequência de divulgação maior. Alguns exemplos de baixa frequência seriam: o PIB, que é divulgado trimestralmente, com um atraso de aproximadamente duas semanas; o IPCA, que é divulgado mensalmente; e a Taxa Selic, que é definida a cada 45 dias etc. Já no trabalho Ghysels et al. (2012), recomenda-se o uso de séries financeiras, que são monitoradas pelo mercado diariamente.

Como referência para o *nowcasting* da inflação usaremos o artigo Bańbura et al. (2013), onde são explorados dados com frequência semanal e diária a fim de produzir estimativas mais precisas de inflação para o mês corrente e meses seguintes. Em particular, este trabalho usou as Estatísticas de Preços do Boletim Semanal do Petróleo para a área do Euro, os preços semanais de gasolina e diesel no varejo para os EUA e o mercado mundial diário de Preços de Matérias-Primas. Seus resultados mostram que esses dados melhoram a precisão das previsões em relação aos modelos que exploram os dados disponíveis apenas com frequência mensal para ambos os países. As evidências empíricas apresentadas por Modugno (2011) mostram que usar dados de alta frequência não só melhoram as previsões para inflação no curto prazo, como também até um ano a frente.

## 2.3 Dilema Viés e Variância

Pensando em avaliar a performance de um modelo de previsão, para um determinado conjunto de dados, é necessário alguma maneira de medir quanto as previsões erram quando comparadas aos dados observados. No método de regressão, a medida mais comumente usada é o erro quadrático médio (Mean Squared Error, MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2, \quad (2.1)$$

onde  $\hat{f}(x_i)$  é a predição que  $\hat{f}$  apresenta para a  $i$ -ésima observação. O MSE será menor ao passo que as respostas previstas forem próximas das respostas verdadeiras. Em aprendizagem estatística, a amostra pode ser separada em mais de um corte, sendo estes cortes divididos em treino e teste, sendo o treino usado para o aprendizado, no intuito de minimizar o erro. O MSE referente ao conjunto de dados de treino dos modelos será aqui chamado como MSE de treino. Em geral não é tão importante saber quão pequeno o MSE possa ser nos dados de treino, mas sim em como o modelo está performando nos dados de teste, que é onde são feitas as primeiras previsões e pode-se inferir se o modelo funciona. Para minimizar o MSE de teste, uma opção é separar uma parte da amostra como um conjunto de dados de teste, ou seja, podemos ter acesso a um conjunto de observações que não foram usadas para treinar o modelo. (GARETH et al., 2013)

Aparentemente o MSE de teste e o MSE de treino possuem alta relação. Todavia, não há garantia que o modelo com o menor MSE de treino também terá o MSE de teste

baixo. O problema de alguns modelos estatísticos é que eles estimam especificamente coeficientes de modo a minimizar o MSE de treino. Para esses modelos, o MSE de treino pode ser pequeno, mas o MSE de teste geralmente é muito maior.

É possível mostrar que o MSE de teste esperado, para um dado valor  $x_0$ , pode sempre ser decomposto na soma de três quantidades fundamentais: a variância de  $\hat{f}(x_0)$ , o viés ao quadrado de  $\hat{f}(x_0)$  e a variância do erro  $\epsilon$ . Pode ser escrito na forma:

$$E \left( y_0 - \hat{f}(x_0) \right)^2 = \text{Var} \left( \hat{f}(x_0) \right) + \left[ \text{Viés} \left( \hat{f}(x_0) \right) \right]^2 + \text{Var}(\epsilon). \quad (2.2)$$

Aqui a notação  $E \left( y_0 - \hat{f}(x_0) \right)^2$  define o MSE de teste esperado em  $(x_0)$  e refere-se ao MSE médio de teste que obteríamos se repetidamente estimasse  $f$  usando um grande número no conjunto de treinamento, e testado cada um em  $(x_0)$ . O MSE de teste geral esperado pode ser calculado pela média  $E \left( y_0 - \hat{f}(x_0) \right)^2$  de todos os valores possíveis de  $x_0$  no conjunto de teste.

A Equação 2.2 mostra que, para minimizar o erro de teste esperado, é necessário selecionar um modelo de aprendizado estatístico que alcance simultaneamente baixa variância e baixo viés. A variância é inerentemente um fator não negativo de quantidade, e o viés quadrado também é não negativo. Assim, o teste MSE esperado nunca pode ficar abaixo de  $\text{Var}(\epsilon)$ , o erro irreduzível.

A variância refere-se à quantidade pela qual  $\hat{f}$  mudaria se o estimássemos usando um conjunto de dados de treinamento diferente. Como os dados de treinamento são usados para ajustar o método de aprendizado estatístico, diferentes conjuntos de dados de treinamento resultam em um  $\hat{f}$  diferente. Mas, idealmente, a estimativa para  $\hat{f}$  não deve variar muito entre as séries de treinamento. No entanto, se um modelo apresentar alta variância então pequenas mudanças nos dados de treinamento podem resultar em grandes mudanças em  $\hat{f}$ . Em geral, modelos mais flexíveis têm maior variância, um modelo mais flexível é aquele que possui um maior número de parâmetros.

Já o viés refere-se ao erro que é introduzido ao relacionar um problema da vida real, o que pode ser extremamente complicado, a um modelo simples. Por exemplo, a regressão linear assume que existe uma relação entre a variável alvo e seus regressores. É improvável que qualquer problema da vida real realmente tenha uma relação linear tão simples e, portanto, realizar a regressão linear resultará em algum viés na estimativa de  $f$ . Geralmente, modelos mais flexíveis resultam em menos viés.

Como regra geral, conforme usamos métodos mais flexíveis, a variância aumenta e o viés diminui. A taxa relativa de variação desses duas quantidades determina se o teste MSE aumenta ou diminui. Como aumentamos a flexibilidade de uma classe de métodos, o viés tende a inicialmente diminuir mais rápido do que a variância aumenta. Consequentemente, o MSE de teste declina. No entanto, em algum ponto, o aumento

da flexibilidade tem pouco impacto no viés, mas começa a aumentar significativamente a variância. Quando isso acontece o MSE de teste aumenta.

O bom desempenho do conjunto de teste de um modelo de aprendizado estatístico requer baixa variância, bem como viés baixo. Isto é referido como um *trade-off*, porque é fácil obter um modelo com viés extremamente baixo, mas alta variância, ou um modelo com variância baixa, mas alto viés. O desafio está em encontrar um método para o qual tanto a variância quanto o viés são baixos.

Seguindo este desafio, a regressão penalizada introduz uma penalização aos regressores (viés), no formato de um  $\lambda$ , que determina quão severa será a penalização, a fim de evitar que o modelo sobreajuste aos dados de treino. Pode não ser imediatamente óbvio porque tal restrição deve melhorar o ajuste, mas acontece que encolher as estimativas dos coeficientes pode reduzir significativamente sua variância. As duas técnicas mais conhecidas para reduzir os coeficientes de regressão em direção a zero são a regressão RIDGE e LASSO, que são explicadas no capítulo 4.

## 3 Dados

Neste trabalho, estamos interessados em previsões semanais, *nowcasting* da inflação corrente, uma vez que é uma das principais métricas macroeconômicas na literatura, sendo uma das variáveis contempladas na maioria das análises macroeconômicas e somente divulgada oficialmente ao término do mês, com atraso de aproximadamente 12 dias. Para isso surge a necessidade de uma ampla base de dados. A base de dados usada para este trabalho foi construída em frequências de amostragem diferentes: uma mensal com preços e indicadores macroeconômicos, uma semanal com preços dos combustíveis e expectativas para o IPCA, e um banco de dados diários com indicadores financeiros.

O ambiente rico em dados da amostra e uma grande quantidade de preditores financeiros, disponíveis em frequência diária, permite-nos realizar o *nowcasting* para o IPCA em qualquer momento do mês, e neste trabalho foi escolhido 4 vezes ao mês. Aqui entra o problema de bordas irregulares (*ragged edge*), porque se em determinado momento do mês, existirem informações ainda não divulgadas, quando realizado o exercício de *nowcasting*, haveria dados faltantes. Isso resulta em um conjunto incompleto de dados para o período mais recente, causando uma borda irregular no final da amostra (WALLIS, 1986). Utilizamos aqui duas soluções: a primeira foi um modelo autoregressivo de primeira ordem (AR1), um modelo de regressão que relaciona uma variável de série temporal a seus valores passados, para os regressores ainda não divulgados. Como esse apresentou resultados aquém do esperado, procuramos outro método, usando para os dados faltantes naquele período o seu último valor observado (modelo do passeio aleatório).

Para garantir estacionariedade dos regressores, os dados são transformados em taxas de crescimento ano a ano ou primeira diferença anual. Isso se aplica tanto para os regressores em baixa frequência quanto aos regressores de alta frequência disponíveis em periodicidade semanal e diária.

Para as séries de dados diários, apresentadas como XD (IPCS e FIPE, FOREX, SELIC, IBOV, IEE, DI, SPREAD, CDS, BCOM), foram criados 4 vetores com os dados relativos ao último lançamento disponível nos dias, 08 (IPCS1, FIPE1), 15 (IPCS2, FIPE2), 22 (IPCS3, FIPE3), e último dia do mês (IPCS4, FIPE4). Como os preços dos ativos são variáveis econômicas voltadas para o futuro, eles constituem uma classe de preditores potencialmente úteis da inflação futura. Stock e Watson (2003).

A fonte de dados para as séries está disponível no apêndice e o calendário de divulgação dos dados tem como fonte a Bloomberg, uma plataforma que coleta as séries de dados de suas fontes originais.

O período escolhido é de 2003 até 2021, sendo o primeiro ano usado para a

construção de índices dos preços, sendo assim, o período amostral é de janeiro de 2004 até dezembro de 2021. Todo o conjunto de dados é dividido em duas partes: a amostra de aprendizagem que vai de 01/01/2004 até 31/12/2012, em que o algoritmo determina qual o melhor valor a se usar para lambda, realizando uma validação cruzada e identificando o valor lambda que produz o menor erro quadrático médio de teste (MSE), e a amostra de teste que inicia em 01/01/2013 e finaliza em 31/12/2021, na qual o algoritmo usa o melhor lambda determinado pelo conjunto de dados teste para gerar previsões e segue se atualizando a cada nova mostra. Mais detalhes são descritos na próxima seção.

### 3.1 Janela Móvel

Usando uma janela móvel, obtemos previsão fora da amostra, iniciando em janeiro de 2003 até dezembro de 2012, assim o modelo “aprende” com os dados quais são os melhores valores estimados dos parâmetros. Para estimar o modelo, usamos esses valores estimados, e movemos a janela rolante para obtermos nossa primeira "previsão", fazendo o modelo voltar ao processo de aprendizado, comparando previsão com o valor observado calculamos o erro de previsão e movermos a janela móvel, gerando uma nova previsão.

### 3.2 Apresentação dos regressores escolhidos

Dados com frequência mensais podem ser observados nas figuras 1 e 2, sendo na tfigura 1 indicadores de preços e a figura 2 indicadores de atividade econômica.

Figura 1 – Dados mensais preço

Series	Mnemonic	Source	Source Units	Relevância e Normalização
(IPCA-15)	IPCA15	IBGE	MoM % change	928.571
National CPI	INPC	IBGE	MoM % change	-
Inflation IGV-M	IGPM	FGV	MoM % change	839.286
Inflation IGV-DI	IGPDI	FGV	MoM % change	660.714
FGV IGP-10	IGP10	FGV	MoM % change	446.429
FGV IPAM	IPAM	FGV	MoM % change	-
FGV IPADI	IPADI	FGV	MoM % change	-
Construct	INCCM	FGV	MoM % change	196.429
FGV CPI	IPCS1	FGV	MoM % change	678.571
IPC-S W1	IPCS2	FGV	MoM % change	678.571
FGV CPI	IPCS3	FGV	MoM % change	678.571
IPC-S W3	IPCS4	FGV	MoM % change	678.571
FGV CPI	FIPE1	Fipe	MoM % change	464.286
IPC-W4	FIPE2	Fipe	MoM % change	464.286
FGV FIPE	FIPE3	Fipe	MoM % change	464.286
CPI W3	FIPE4	Fipe	MoM % change	464.286
FGV FIPE				
CPI W4				

Fonte: Autor (2022).

Indicadores de atividade econômica mensais

Figura 2 – Dados mensais atividade econômica

Economic Activity					
18	Industrial Production	IP	IBGE	MoM % change	821.429
19	Economic Activity	IBCB	Bacen	MoM % change	910.714
20	Activity Index	IBCB	Bacen	MoM % change	607.143
21	Brazil PMI	PMIM	IHS Markit	MoM % change	553.571
22	Brazil PMI	PMIS	IHS Markit	MoM % change (threshold = 50)	90
23	Retail Sales	RETSAL	IBGE	MoM % change	785.714
24	Services Volume	SERVOL	IBGE	MoM % change	892.857
25	FGV Consumer Confidence	CONCONF	FGV	Index	178.571
26	CNI Business Confidence	BUSCONF	CNI	Index (threshold = 50)	267.857
27	Unemployment	UNEMP	IBGE	Percent of labor force	-
28	Payroll	JOB	MTb/CAGED	Units	285.714
29	Vehicle Sales	VEHSALE	Fenabrave	Units	767.857
30	Vehicle Sales	VEHSALE	Fenabrave	Units	428.571
	Vehicle Sales	VEHSALE	Fenabrave	Units	107.143

Fonte: Autor (2022).

Os Dados de frequência semanal podem ser observados nas figuras 3 e 4.

Figura 3 – Dados semanais combustíveis

Series	Mnemonic	Source	Source Units	Relevance	
48	IPCA Expectations t+0	EXPT0	Bacen	MoM % change	-
49	IPCA Expectations t+1	EXPT1	Bacen	MoM % change	-
50	IPCA Expectations t+2	EXPT2	Bacen	MoM % change	-
51	IPCA Expectations t+3	EXPT3	Bacen	MoM % change	-

Fonte: Autor (2022).

Figura 4 – Dados semanais expect

Series	Mnemonic	Source	Source Units	Relevance	
Weekly Data					
33	Diesel	DIESEL	ANP	R\$/Liter	-
34	Regular Gasoline	GAS	ANP	R\$/Liter	-
35	Ethanol Fuel	ETOH	ANP	R\$/Liter	-
36	Liquefied Natural Gas	LNG	ANP	R\$/13Kg	-

Fonte: Autor (2022).

E os dados de frequência diária que podem ser observados na figura 5.

Figura 5 – Dados de frequência diária

	1	2	3	4	5
1	Series	Mnemonic	Source	Source Units	Relevanc e Bloombe
37	<b>Daily Data</b>				
38					
	Selic Interest Rate	SELIC	Bacen	% p.y.	982.143
39	Exchange Rate - BR Real/US\$	FOREX	Bacen	Level	-
40	Stock Price Index: IBOV	IBOV	BM&FBovespa	Index	-
41	Electric Utilities Index	IEE	BM&FBovespa	Index	-
42	DI rates 10-Year	DI10	BM&FBovespa	% p.y.	-
43	Maturity Spread of Swap DI rates (10-Brazil)	SPREAD	BM&FBovespa	% p.y.	-
44	Credit Default Bloomborg	CDS	BM&FBovespa	Basis points	-
45	Bloomborg	BCOM	Bloomborg	Index	-
46	Commod				

Fonte: Autor (2022).

# 4 Metodologia

## 4.1 Regressão Linear

Considerando o modelo padrão de mínimos quadrados ordinários (MQO) para regressão linear múltipla:

$$Y = X\beta + \epsilon, \quad (4.1)$$

onde  $y \in \mathbb{R}^n$ ,  $\beta \in \mathbb{R}^p$ , e  $X \in \mathbb{R}^{n \times p}$ . Podemos expandir isso para

$$y_i = \sum_{j=1}^p \beta_j X_{ij} + \epsilon_i \forall i = 0, 1, \dots, n. \quad (4.2)$$

Aqui  $\beta_j$  são parâmetros desconhecidos não aleatórios,  $X_{ij}$  são observáveis, e  $\epsilon_i$  são aleatórios, então  $y_i$  são aleatórios. O estimador de MQO escolhe os coeficientes de regressão de modo que a reta de regressão estimada seja a mais próxima possível dos dados observados, em que a proximidade é medida pela soma dos quadrados dos erros de previsão de  $y$  dado  $X$ . (STOCK; WATSON, 2012)

### 4.1.1 Multicolinearidade e Sobreajuste

Um dos problemas mais comuns com o método MQO é a tendência do modelo a sobreajustar os dados quando há muito ruído causado por variáveis correlacionadas. Isso pode acontecer em muitas situações diferentes. O caso mais extremo ocorre quando  $p > n$ . Da regressão linear múltipla temos a estimativa do coeficiente

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4.3)$$

que podemos reescrever como  $[(X^T X)^{-1} X^T]^{-1} \hat{\beta} = Y$ . Podemos ver que se  $p > n$  não existe uma solução única para o sistema a matriz  $X$  passa a possuir mais colunas do que linhas, e quando isso acontece,  $(X^T X)$  não é inversível, não havendo uma matriz de posto completo, não existe uma solução única para os mínimos quadrados dos resíduos e a regressão linear não produz valores de coeficiente precisos. Com situações menos extremas, a multicolinearidade pode fazer com que o modelo seja excessivamente sensível a pequenas mudanças nos valores dos parâmetros e os coeficientes podem ter o sinal errado ou uma ordem de grandeza incorreta. Quando um modelo linear começa a sobreajustar os dados, os coeficientes podem apresentar altos erros padrão e baixos níveis de significância, apesar de um alto valor de  $R^2$ .

## 4.2 Regressão RIDGE, LASSO e Elastic Net

### 4.2.1 Introdução à regularização

A regularização é um método para resolver problemas de multicolinearidade ou problemas de modelos que sobreajustam os dados. O método envolve a introdução de informações adicionais a um modelo na forma de uma penalização. Em termos de Ridge e LASSO, a penalidade impõe um encolhimento nas estimativas dos coeficientes dos mínimos quadrados ordinários. Essa penalidade controla a instabilidade encontrada no modelo de mínimos quadrados com matrizes não ortogonais. Geralmente, para o termo de regularização  $L_p$  temos  $L_p = (\sum_i \|\beta_i\|^p)^{\frac{1}{p}}$ . Ridge e LASSO lidam com as penalidades de  $L_2$  e  $L_1$ , respectivamente. A regularização é usada preferencialmente em relação a outros métodos comuns para determinar o melhor modelo linear, como a seleção do melhor subconjunto e a seleção gradual do subconjunto (DUSEN, 2016).

### 4.2.2 Regressão Ridge

Dada a soma da estimativa de erro quadrado para mínimos quadrados, temos  $(Y - X\beta)^T(Y - X\beta)$ . A regressão RIDGE adiciona a penalidade  $L_2$  de tal forma que temos

$$(Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta \quad (4.4)$$

Da equação (4.3) podemos derivar a estimativa do coeficiente RIDGE:

$$\begin{aligned} & (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta \\ &= (Y^T - \beta^T X^T)(Y - X\beta) + \lambda\beta^T\beta \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta + \lambda\beta^T\beta \\ &= Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta + \lambda\beta^T\beta \\ &\rightarrow \frac{d}{d\beta} = 0 - 2Y^T X + 2X^T X\hat{\beta} + 2\lambda\beta = 0 \\ &\rightarrow \hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T Y \end{aligned} \quad (4.5)$$

Nas equações (4.3) e (4.4),  $\lambda \geq 0$  é um parâmetro de ajuste para a penalização, que é determinada separadamente. Quando  $\lambda = 0$  temos a estimativa dos mínimos quadrados ordinários, e quando  $\lambda \rightarrow \infty$  todos os coeficientes se aproximam de zero. À medida que  $\lambda$  aumenta, a flexibilidade do ajuste do modelo diminui. Isso leva ao aumento do viés, mas à diminuição da variância. Enquanto RIDGE reduz os coeficientes para zero, o modelo final escolhido sempre incluirá todos os regressores (a menos que  $\lambda = \infty$ , caso em que todos os regressores serão zero).

### 4.2.3 LASSO

Podemos demonstrar o modelo LASSO da mesma forma que a equação (4.4) acima:

$$(Y - X\beta)^T(Y - X\beta) + \lambda|\beta|_1 \quad (4.6)$$

Onde  $|\beta|_1 = \sum_{j=1}^p |\beta|_j$ . A diferença é que LASSO usa a penalidade  $L_1$  em vez da penalidade  $L_2$ . O maior benefício do LASSO é a capacidade do modelo de criar matrizes esparsas (uma grande quantidade de elementos com valor 0). A grande diferença entre LASSO e RIDGE é que RIDGE encolhe parâmetros para valores próximos de zero, enquanto LASSO pode reduzir os parâmetros até zero, fazendo com que esses desapareçam.

### 4.2.4 Elastic Net

*Elastic Net* aplica uma combinação das penalidades  $L_1$  e  $L_2$

$$p_{\lambda_1, \lambda_2}(|\beta|) = \lambda_1|\beta| + \lambda_2\beta^2. \quad (4.7)$$

Como o LASSO, a Elastic Net realiza a seleção automática de variáveis definindo algumas estimativas de coeficiente como zero. Mas o termo adicional de penalidade  $L_2$  distribui o peso para mais variáveis, de modo que a Elastic net tende a selecionar mais variáveis do que o LASSO. Esse é especialmente o caso em situações com alta correlação, nos quais o LASSO selecionaria apenas uma variável de um conjunto de variáveis altamente correlacionadas, enquanto a penalidade do RIDGE lhes daria peso igual e esses parâmetros apenas teriam valores muito baixo.

## 4.3 Avaliação de performance do modelo

Para avaliar a acurácia, usaremos a raiz quadrática média dos erros (RMSE) entre valores observados (reais) e predições (hipóteses).

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

onde  $N = t^{(m)}$ , sendo  $t$  o numero de observações da inflação e  $m = 4$ ,  $\hat{y}_i$  é o valor previsto para a observação no conjunto de dados,  $y_i$  é o valor observado no conjunto de dados.

Este resultado é importante, porque ele diz quanto o modelo está errando em relação a inflação observada efetivamente. O RMSE relativo ao passeio aleatório é usado como base para a criação de um índice, e assim avaliar se realizar o *nowcasting* traz maior poder preditivo que simplesmente assumir que a inflação se manteria onde estava na ultima observação.

## 4.4 Modelo de frequência mista irrestrito

Considerando que as variáveis mensais de baixa frequência  $y_t$  e  $x_t$  estejam associadas a índices de tempo  $t = 1, \dots, T$  enquanto os preditores de alta frequência  $x_t^{(m)}$  são amostrados  $m$  vezes com mais frequência. A partir daí, será conveniente converter as variáveis financeiras diárias para a taxa de amostragem dos preditores semanais e, por duas razões, assumimos uma frequência fixa mês/semana de  $m = 4$ . Primeiro, foi escolhido uma mistura de apenas duas frequências para evitar uma perda de qualidade decorrente de uma maior incompatibilidade de frequência mensal/diária. Segundo, alguns dos ruídos e volatilidade diários, exibidos por algumas das séries financeiras diárias, são suavizados.

Para este fim, os preditores amostrados em frequências mais altas serão representados pelas informações mais recentes disponíveis nos dias 8, 15, 22 e no final de cada mês, de modo que o alinhamento da frequência com a variável de destino  $y_t$  seja alcançado. Isso significa que usamos apenas as quatro observações submensais recentes para obter o *nowcasting*.<sup>1</sup> Portanto, para cada realização na frequência mais baixa  $t$ , o conjunto de informações  $\Omega_t$  também inclui observações de alta frequência submensais:  $x_t^{(m)}, x_{t-\frac{1}{m}}^{(m)}, x_{t-\frac{2}{m}}^{(m)}, x_{t-\frac{m-1}{m}}^{(m)}$ , onde  $t - k/m$  denota o  $k^{\text{th}}$  período de alta frequência passado com  $k = 0, \dots, m - 1$ .<sup>2</sup> Em particular, as observações de fim de mês correspondem a  $k = 0$ .

Dado um conjunto de preditores de alta frequência  $x_t^{(m)} = (x_{1t}^{(m)}, \dots, x_{Nt}^{(m)})$  e preditores de baixa frequência  $x_t = (x_{1t}, \dots, x_{Rt})$ , o modelo de previsão de frequência mista mais geral para  $y_{t+h}$ , condicionado à dinâmica recente de  $x_t^{(m)}$  e comportamento passado de  $y_t$  e  $x_t$ , é explicitamente dado por:

$$y_{t+h} = c + \sum_{j=1}^p \rho_j y_{t+1-j} + \sum_{n=1}^N \sum_{k=0}^K b_{nk} x_{n,t+1-\frac{k}{m}}^{(m)} + \sum_{n=1}^R \sum_{j=1}^q \alpha_{nj} x_{n,t+1-j} + \varepsilon_{t+h}, \quad (4.8)$$

onde  $K$  é o número de defasagens distribuídas na frequência de amostragem de  $x_t^{(m)}$  e  $\varepsilon_{t+h}$  são erros normalmente distribuídos. Modelo (4.8) pode ser reescrito de forma compacta como:

$$A(L) y_{t+h} = c + \sum_{n=1}^N B(L^{1/m}) x_{n,t+1}^{(m)} + \sum_{r=1}^R Z(L) x_{r,t+1} + \varepsilon_{t+h}, \quad (4.9)$$

onde  $B(L^{1/m}) = \sum_{k=0}^K b_{nk} L^{k/m}$  é um polinômio de comprimento  $(K + 1)$  no  $L^{1/m}$  operador com  $L^{k/m} x_{n,t}^{(m)} = x_{n,t-k/m}^{(m)}$ . Apesar de  $\Omega_t$  poder incluir atrasos de alta frequência  $K$  em vários períodos anteriores de baixa frequência, assumimos  $K = m - 1$  para simplificar e, assim, o modelo (4.9) se torna estático no componente de alta frequência. Isso significa que usamos apenas as quatro observações submensais recentes para obter o *nowcasting*.

<sup>1</sup> Observe que essa escolha específica de dias permite controlar os problemas de semanas sobrepostas em meses e número irregular de dias úteis ou semanas em meses diferentes.

<sup>2</sup> Observe que o tempo inteiro  $t$  ainda está na variável de baixa frequência porque as variáveis de alta frequência são observadas em  $1/m$  frações dentro de cada período  $t$ .

Vale ressaltar que o modelo (4.9) envolve uma série de regressores  $d = (KN + qR + p + 1)$  e, conseqüentemente, a estratégia de defasagens distribuídas irrestritas pode facilmente levar à proliferação de parâmetros, conforme aumento do número de preditores de alta frequência  $N$  e atrasos  $K$ . Embora a estimativa MQO direta seja ótima na configuração linear (4.9), ela anda de mãos dadas com uma incerteza de estimativa muito alta quando o número de parâmetros  $d$  é muito grande em relação ao tamanho da amostra  $T$ . Para superar esse problema de Big Data, empregamos os métodos de regularização/encolhimento, que são descritos em detalhes acima.

#### 4.4.1 Modelo de frequência mista irrestrito em notação matricial

Por simplicidade expositiva, vamos reduzir o caso geral de preditores múltiplos na Eq. (4.9) para o caso de um preditor, tanto de alta quanto de baixa frequência, e definindo  $p = q = 1$  enquanto mantém  $m = 4$ . A partir daí, condicionalmente aos dados de alta frequência disponíveis até o período  $T$ , o modelo (4.9) pode ser estimado seguindo a notação matricial

$$\begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & y_1 & x_2^{(m)} & x_{2-1/m}^{(m)} & x_{2-2/m}^{(m)} & x_{2-3/m}^{(m)} & x_1 \\ 1 & y_2 & x_3^{(m)} & x_{3-1/m}^{(m)} & x_{3-2/m}^{(m)} & x_{3-3/m}^{(m)} & x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{T-1} & \underbrace{x_T^{(m)}}_{\text{end of month}} & \underbrace{x_{T-1/m}^{(m)}}_{\text{day 22}} & \underbrace{x_{T-2/m}^{(m)}}_{\text{day 15}} & \underbrace{x_{T-3/m}^{(m)}}_{\text{day 8}} & x_{T-1} \end{bmatrix} \begin{bmatrix} c \\ \rho_1 \\ b_{11} \\ b_{12} \\ b_{13} \\ b_{14} \\ \alpha_{11} \end{bmatrix} + \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix} \quad (4.10)$$

Observe que a representação matricial (4.10) torna explícita a transformação das variáveis de alta frequência  $x_{t-k/m}^{(m)}$  em um vetor de baixa frequência  $(x_{p+1-k/m}^{(m)}, \dots, x_{T-k/m}^{(m)})'$ , para  $k = 0, \dots, m - 1$ . Dessa forma o modelo (4.9) é estimado na baixa frequência, enquanto o alinhamento de frequência requer que o número de observações de cada preditor de alta frequência  $x_t^{(m)}$  seja igual a  $mT$  tal que possa ser decomposto em incrementos de  $m$  dentro de cada período  $t$ . Além disso, é possível reestimar essa representação em cada ponto no tempo escolhido submensal à medida que os dados de alta frequência se tornam disponíveis.

# 5 Resultados e discussão

## 5.1 Comparando a performance dos modelos

Para comparar o desempenho dos modelos descritos acima na tarefa de *nowcasting*, os erros quadráticos médios foram calculados para os resultados de cada modelo. Para evidenciar os benefícios do uso dos modelos em vez do método do passeio aleatório, as tabelas relatam o erro de previsão de cada modelo em relação ao erro de previsão do modelo de passeio aleatório. Para fim de facilitar o entendimento do leitor, foi criado um índice, usando como base o passeio aleatório, sendo assim, qualquer valor final abaixo de 100 representa uma melhoria relativa no modelo, de escala igual a diferença entre eles, e um valor acima implicaria em uma inferioridade. Também foi incluído o cálculo dos erros quadráticos médios das expectativas de mercado do boletim Focus.

### 5.1.1 Solução Autoregressiva para *ragged edge*

Como mencionado no capítulo 3, uma das metodologias adotadas para o problema de *ragged edge* é a substituição dos dados faltantes por uma previsão autoregressiva do mesmo.

A Tabela 1 relata o erro de previsão de cada modelo em relação ao erro de previsão do Modelo de passeio aleatório: erro do modelo/erro RW, os indicadores ( $w_1$ ,  $w_2$ ,  $w_3$ ,  $w_4$ ) se referem as 4 observações mensais realizadas nos dias (08,15,22, e último dia do mês)

#### 5.1.1.1 Método dos mínimos quadrados

O modelo MQO apresentou uma diferença de +19 pontos percentuais em relação ao modelo de random walk, implicando em uma piora na previsão frente ao RW. As semanas 2 e 3 mostram um erro muito expressivo.

#### 5.1.1.2 Método de regularização RIDGE

Realizando *nowcasting* com RIDGE, obtivemos uma melhoria de 29,3 pontos percentuais em relação ao modelo de passeio aleatório, portanto, a seleção de variáveis feita traz aumento da qualidade de previsão da inflação. É possível observar que ao longo das semanas, conforme as informações são adicionadas ao modelo, seu poder preditivo passa a aumentar, chegando a uma melhoria de 43 pontos percentuais quando realizado o *nowcasting* no último dia do mês.

### 5.1.1.3 Método de regularização LASSO

Realizando *nowcasting* com LASSO, obtivemos uma melhoria de 25,2 pontos percentuais em relação ao modelo de passeio aleatório, contudo, sendo menos indicada aqui do que o RIDGE, possivelmente o modelo encolheu alguns parâmetros até zero, perdendo um pouco do poder de previsão. É possível observar que ao longo das semanas, conforme as informações são adicionadas ao modelo, seu poder preditivo passa a aumentar, chegando a uma melhoria de 46 pontos percentuais quando realizado o *nowcasting* no último dia do mês.

### 5.1.1.4 Método de regularização Elastic Net

O Elastic Net trouxe uma melhoria de 30,6 pontos percentuais em relação ao passeio aleatório, apresentou maior poder de previsão frente ao RIDGE e LASSO, o que faz sentido, tendo em vista que ele realiza a validação cruzada e escolhe entre L1 e L2. É possível observar que ao longo das semanas, conforme as informações são adicionadas ao modelo, seu poder preditivo passa a aumentar, chegando a uma melhoria de 48 pontos percentuais quando realizado o *nowcasting* no último dia do mês.

### 5.1.1.5 Expectativas do mercado Boletim Focus

As expectativas de mercado apresentam uma melhoria em relação ao passeio aleatório de 60,3 pontos percentuais, tendo performado muito melhor que o nosso modelo. Este fato nos fez buscar outra solução para o problema de *ragged edge*, e o resultado é apresentado abaixo. As expectativas também melhoram seu poder preditivo ao longo do mês, não tão expressiva quanto nosso modelo, mas mantendo uma lógica e chegando a 62 pontos percentuais.

## 5.1.2 Solução de passeio aleatório para *ragged edge*

Como mencionado no capítulo 3, uma das metodologias adotadas para o problema de *ragged edge* é a substituição dos dados faltantes por uma previsão de passeio aleatório do mesmo.

A Tabela 2 relata o erro de previsão de cada modelo em relação ao erro de previsão do Modelo de passeio aleatório: erro do modelo/erro RW, os indicadores ( $w_1$ ,  $w_2$ ,  $w_3$ ,  $w_4$ ) se referem as 4 observações mensais realizadas nos dias (08,15,22, e último dia do mês)

Como este método nos trouxe melhores resultados, apresentaremos também mapas de calor dos modelos de regularização para facilitar a visualização de quais regressores são mantidos, em qual período e relevância.

Tabela 1 – RMSE solução AR para Raged Edge

Modelo	RMSE	RMSE/RMSERW	
RW_RMSE_Nowc_w1	0,83144	100	
RW_RMSE_Nowc_w2	0,47734	100	100
RW_RMSE_Nowc_w3	0,47734	100	
RW_RMSE_Nowc_w4	0,47734	100	
MQO_AR_RMSE_Nowc_w1	0,40121	48,25483499	
MQO_AR_RMSE_Nowc_w2	0,59165	123,9472912	<b>119</b>
MQO_AR_RMSE_Nowc_w3	1,09206	228,7803243	
MQO_AR_RMSE_Nowc_w4	0,35382	74,12326643	
RIDGE_AR_RMSE_Nowc_w1	0,47488	57,11536611	
RIDGE_AR_RMSE_Nowc_w2	0,44362	92,93585285	<b>70,7</b>
RIDGE_AR_RMSE_Nowc_w3	0,35886	75,17911761	
RIDGE_AR_RMSE_Nowc_w4	0,27552	57,71986425	
LASSO_AR_RMSE_Nowc_w1	0,49847	59,95261234	
LASSO_AR_RMSE_Nowc_w2	0,50329	105,4363766	<b>74,8</b>
LASSO_AR_RMSE_Nowc_w3	0,38183	79,99120124	
LASSO_AR_RMSE_Nowc_w4	0,25713	53,86726442	
E.NET_AR_RMSE_Nowc_w1	0,46223	55,59390936	
E.NET_AR_RMSE_Nowc_w2	0,45403	95,11668831	<b>69,4</b>
E.NET_AR_RMSE_Nowc_w3	0,35389	74,13793103	
E.NET_AR_RMSE_Nowc_w4	0,25086	52,55373528	
RMSE_Expec_w1	0,26076	31,3624555	
RMSE_Expec_w2	0,22858	47,88620271	<b>39,7</b>
RMSE_Expec_w3	0,19587	41,03364478	
RMSE_Expec_w4	0,18295	38,32697867	

### 5.1.2.1 Método dos mínimos quadrados

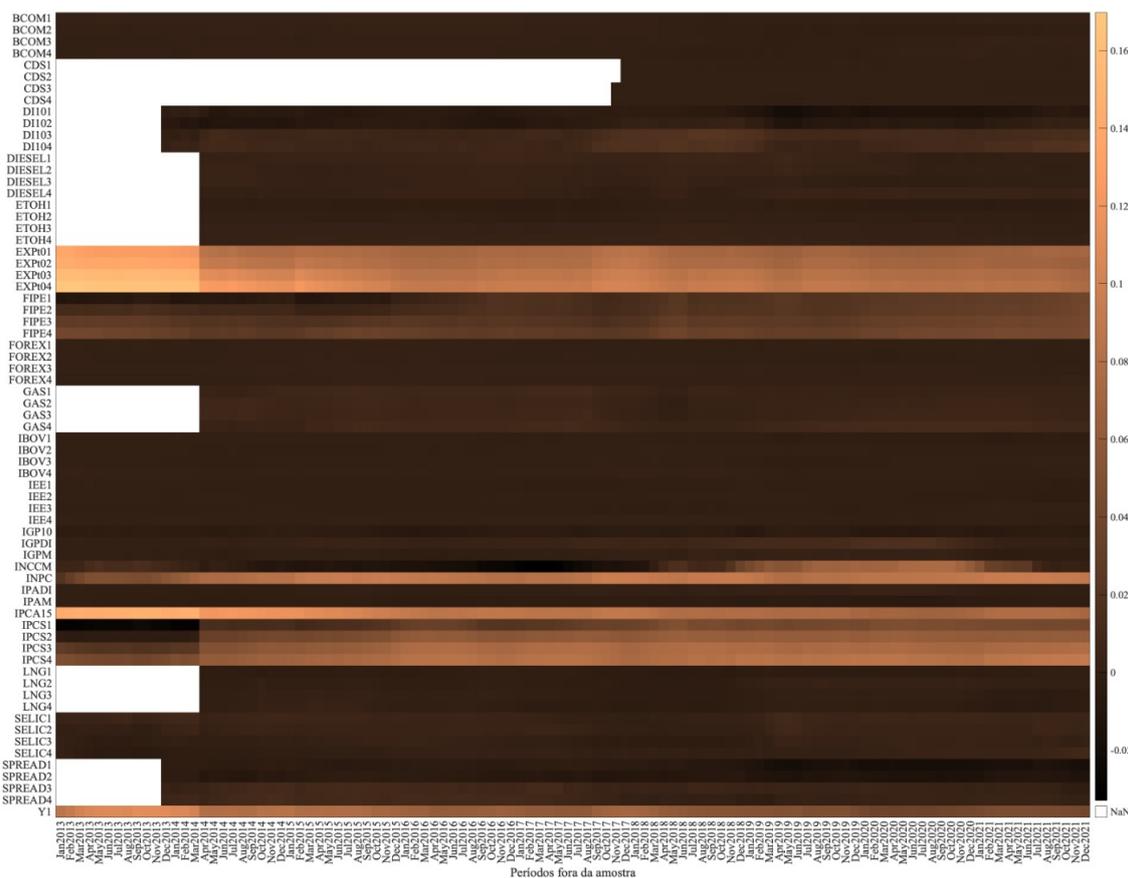
O modelo MQO apresentou uma diferença de 19,1 pontos percentuais em relação ao modelo de random walk, implicando em uma melhora na previsão frente ao RW. É possível observar que ao longo das semanas, conforme as informações são adicionadas ao modelo, seu poder preditivo passa a aumentar, chegando a uma melhoria de 56 pontos percentuais quando realizado o *nowcasting* no último dia do mês.

### 5.1.2.2 Método de regularização RIDGE

Realizando *nowcasting* com RIDGE, obtivemos uma melhoria de 41,4 pontos percentuais em relação ao modelo de passeio aleatório, portanto, a seleção de variáveis feita traz aumento da qualidade de previsão da inflação. É possível observar que ao longo

das semanas, conforme as informações são adicionadas ao modelo, seu poder preditivo passa a aumentar, chegando a uma melhoria de 50 pontos percentuais quando realizado o *nowcasting* no último dia do mês.

Figura 6 – Gráfico de calor RIDGE



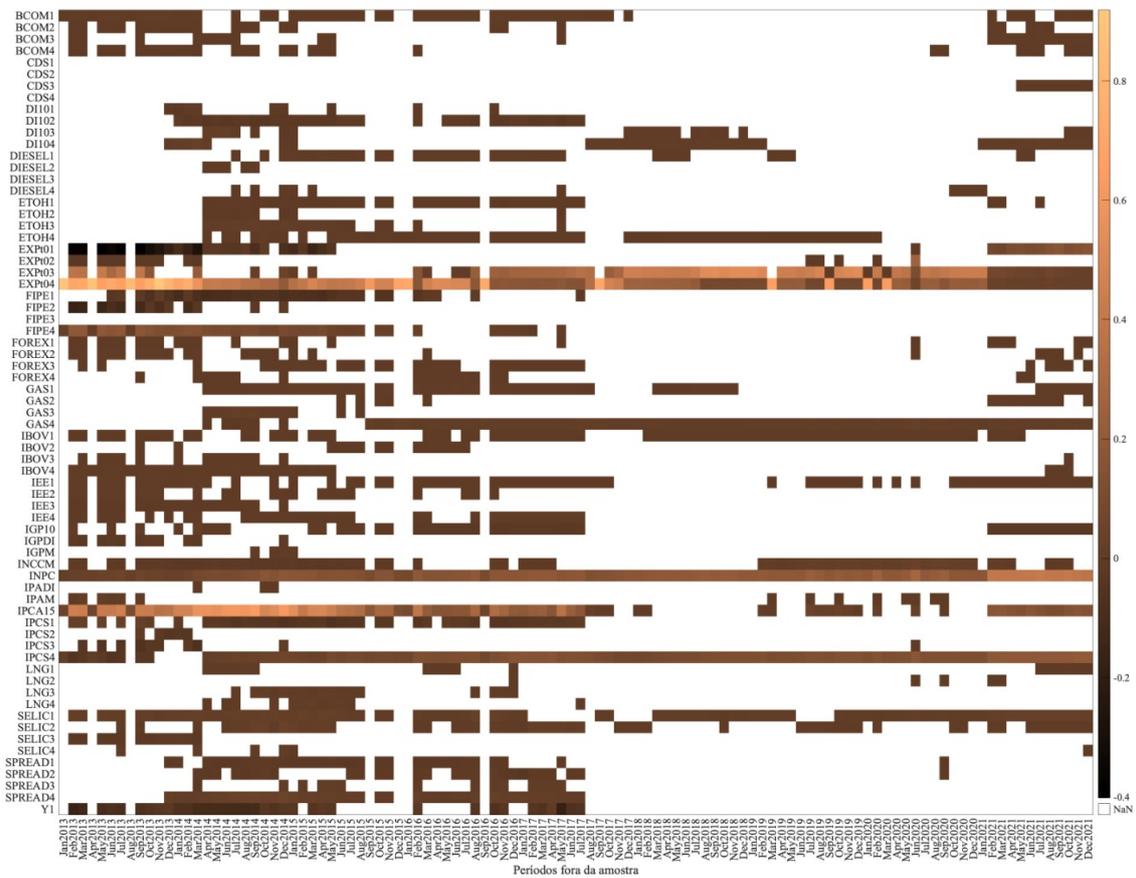
Fonte: Autor (2022).

A figura 1 nos mostra que o Ridge selecionou variáveis e basicamente manteve sua relevância em toda a amostra. Os momentos em branco, que são iguais a zero, correspondem aquelas variáveis que ainda não entraram no painel balanceado por falta de dados no início da amostra.

### 5.1.2.3 Método de regularização LASSO

Realizando *nowcasting* com LASSO, obtivemos uma melhoria de 41,5 pontos percentuais em relação ao modelo de passeio aleatório, apenas 0,1 melhor que o ridge. É possível observar que ao longo das semanas, conforme as informações são adicionadas ao modelo, seu poder preditivo passa a aumentar, chegando a uma melhoria de 55 pontos percentuais quando realizado o *nowcasting* no último dia do mês.

Figura 7 – Gráfico de calor LASSO



Fonte: Autor (2022).

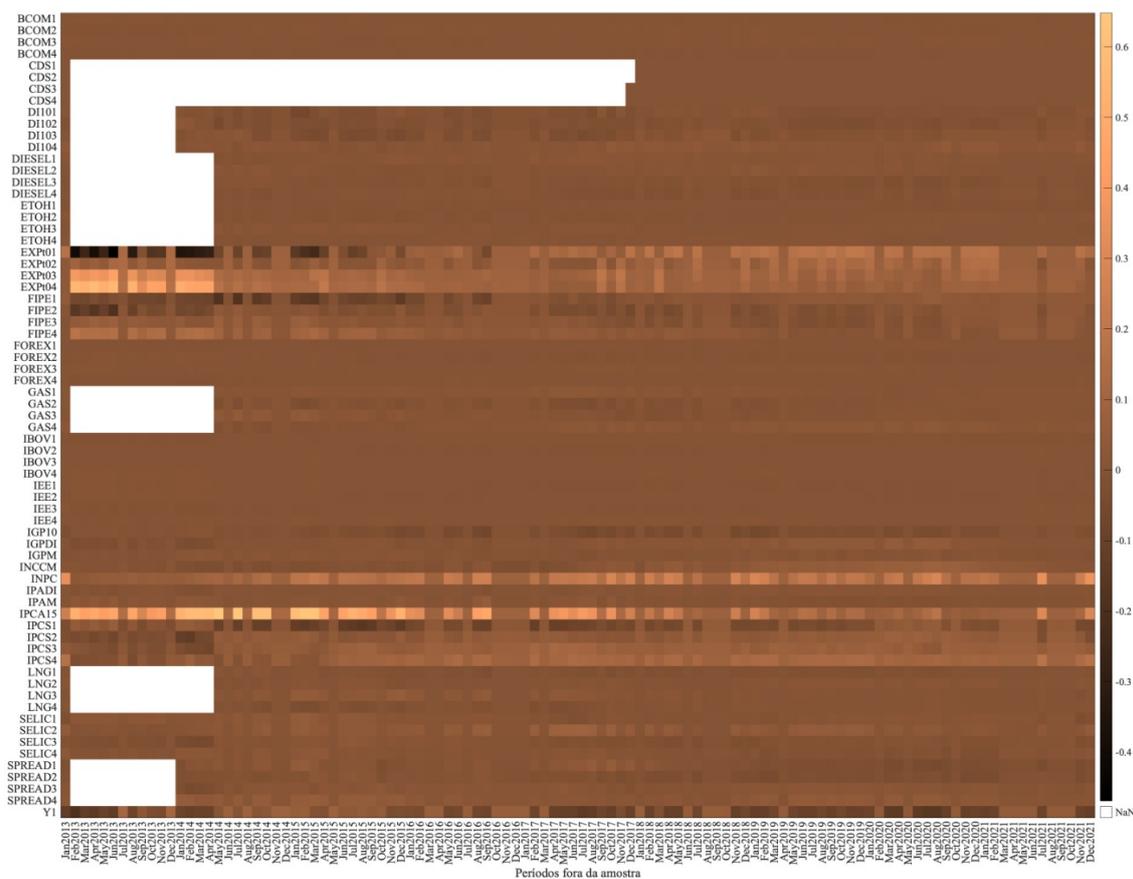
A figura 2 nos mostra que o LASSO alternou entre seus regressores, dando relevância a diferentes, em diferentes épocas. Ele Selecionou as variáveis que tem a informação relevante e descartou as redundâncias. É interessante que alguns regressores se mostraram fortes e foram reduzindo ao longo do tempo.

#### 5.1.2.4 Método de regularização Elastic Net

O Elastic Net trouxe uma melhoria de 30,8 pontos percentuais em relação ao passeio aleatório, apresentou menor poder de previsão frente ao RIDGE e LASSO, o que faz não sentido, tendo em vista que ele realiza a validação cruzada e escolhe entre L1 e L2, sendo assim deveria se manter próximo ou igual aos dois.

Como é possível observar na figura 3, ele escolheu os alfas ótimos de forma a aproximar o modelo de RIDGE, não eliminando muitos parâmetros e mantendo certa constância de relevância a eles.

Figura 8 – Gráfico de calor Elastic Net



Fonte: Autor (2022).

### 5.1.2.5 Expectativas Boletim Focus

As expectativas de mercado apresentam uma melhoria em relação ao passeio aleatório de 60,3 pontos percentuais, tendo performado muito melhor que o nosso modelo. As expectativas também melhoram seu poder preditivo ao longo do mês, não tão expressiva quanto nosso modelo, mas mantendo uma lógica e chegando a 62 pontos percentuais.

Tabela 2 – RMSE solução RW para Raged Edge

Modelo	RMSE	RMSE/RMSERW	
RW_RMSE_Nowc_w1	0,83144	100	
RW_RMSE_Nowc_w2	0,47734	100	
RW_RMSE_Nowc_w3	0,47734	100	
RW_RMSE_Nowc_w4	0,47734	100	
MQO_RW_RMSE_Nowc_w1	0,43383	52,17814875	
MQO_RW_RMSE_Nowc_w2	0,50888	106,6074496	<b>78,9</b>
MQO_RW_RMSE_Nowc_w3	0,54	113,1269116	
MQO_RW_RMSE_Nowc_w4	0,20856	43,6921272	
RIDGE_RW_RMSE_Nowc_w1	0,43191	51,94722409	
RIDGE_RW_RMSE_Nowc_w2	0,34953	73,22453597	<b>58,6</b>
RIDGE_RW_RMSE_Nowc_w3	0,28231	59,14233041	
RIDGE_RW_RMSE_Nowc_w4	0,2384	49,94343654	
LASSO_RW_RMSE_Nowc_w1	0,49511	59,54849418	
LASSO_RW_RMSE_Nowc_w2	0,33924	71,06883982	<b>58,5</b>
LASSO_RW_RMSE_Nowc_w3	0,27919	58,48870826	
LASSO_RW_RMSE_Nowc_w4	0,21442	44,91976369	
E.NET_RW_RMSE_Nowc_w1	0,46206	55,57346291	
E.NET_RW_RMSE_Nowc_w2	0,4575	95,84363347	<b>69,2</b>
E.NET_RW_RMSE_Nowc_w3	0,35391	74,14212092	
E.NETRW_RMSE_Nowc_w4	0,2453	51,38894708	
RMSE_Expec_w1	0,26076	31,3624555	
RMSE_Expec_w2	0,22858	47,88620271	<b>39,7</b>
RMSE_Expec_w3	0,19587	41,03364478	
RMSE_Expec_w4	0,18295	38,32697867	

## 6 Conclusão

Este trabalho apresentou resultados empíricos originários do *nowcasting* da inflação brasileira usando técnicas simples e de aprendizado de máquina. Primeiro coletamos um vasto conjunto de dados com frequência mista e os adaptamos para uma taxa fixa de 4 semanas no mês. Em seguida, um modelo de MQO e 3 modelos de Machine Learning foram aprendidos na amostra de 01/01/2004 a 31/12/2012 e testados na amostra de 01/01/2013 a 31/12/2021. Sua qualidade preditiva foi avaliada pelo RMSE, calculado a partir do erro de previsão em relação ao valor observado. As três tarefas que foram declaradas na introdução como os objetivos deste trabalho foram realizadas, e seus resultados podem ser resumidos da seguinte forma:

(1) A base de dados construída para esse trabalho foi formada por 34 séries de tempo, sendo algumas delas divididas em 4 séries ao longo do mês, somando assim 69 séries macroeconômicas com 216 observações cada;

(2) Há um benefício em usar técnicas de aprendizado de máquina relativamente a um passeio aleatório na previsão da inflação que variam de 5%, podendo chegar a 45% em determinados momentos do mês, chegando muito próximo as expectativas de mercado. Já no valor agregado, os modelos trazem uma melhoria de 31,8% a 42%;

(3) Os modelos apresentam grande utilidade e poder preditivo para o caso da inflação brasileira em relação a um passeio aleatório, todavia quando comparados as expectativas de mercado os modelos chegam próximos mas não melhor.

Este trabalho trata de medir os benefícios do *nowcasting*; Todavia, o esforço para realizar esta tarefa deve também ser considerado. Organizar a base de dados exige tempo e manipular-la exige conhecimento, parece que o caminho mais importante para pesquisas futuras neste ramo é a otimização dos códigos de computador que realizam *nowcasting*, para que esses possam organizar os dados de forma automatizada, tendo em vista que o *nowcasting* da inflação é uma tarefa que impõe restrições de tempo na velocidade dos resultados. Se um humano tiver um custo de completar esta tarefa em alguns dias, o exercício de previsão se torna sem sentido, porque até então a verdadeira taxa de inflação se torna conhecida.

Superando esse problema, podemos sugerir melhorias futuras com a introdução de outros métodos de *nowcasting*, e até mesmo revendo algumas soluções aqui usadas. Um exemplo disso é o comparativo entre as soluções para o *ragged edge*, que nos trouxeram melhorias significativas no poder de previsão.

# Referências

BABII, A.; GHYSELS, E.; STRIAUKAS, J. Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, Taylor & Francis, p. 1–23, 2021.

BAÑBURA, M.; GIANNONE, D.; MODUGNO, M.; REICHLIN, L. Now-casting and the real-time data flow. In: *Handbook of economic forecasting*. [S.l.]: Elsevier, 2013. v. 2, p. 195–237.

DUSEN, C. V. Methods to prevent overwriting and solve ill-posed problems in statistics: Ridge regression and lasso. *Preprint submitted to Colorado College Department of Mathematics September*, v. 16, 2016.

GARETH, J.; DANIELA, W.; TREVOR, H.; ROBERT, T. *An introduction to statistical learning: with applications in R*. [S.l.]: Springer, 2013.

GHYSELS, E.; KOURTELLOS, A.; ANDREOU, E. et al. Should macroeconomic forecasters use daily financial data and how? In: SOCIETY FOR ECONOMIC DYNAMICS. *2012 Meeting Papers*. [S.l.], 2012.

GIANNONE, D.; REICHLIN, L.; SMALL, D. Nowcasting: The real-time informational content of macroeconomic data. *Journal of monetary economics*, Elsevier, v. 55, n. 4, p. 665–676, 2008.

STOCK, J.; WATSON, M. *Introducción a la econometría (3ª Edición ed.)*. [S.l.]: Madrid: Pearson Education, SA, 2012.

STOCK, J. H.; WATSON, M. W. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, v. 41, n. 3, p. 788–829, 2003.

STOCK, J. H.; WATSON, M. W. Phillips curve inflation forecasts. National Bureau of Economic Research, 2008.

WALLIS, K. F. Forecasting with an econometric model: The ‘ragged edge’ problem. *Journal of Forecasting*, Wiley Online Library, v. 5, n. 1, p. 1–13, 1986.