



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE ECONOMIA PROFISSIONAL
MESTRADO PROFISSIONAL EM ECONOMIA DE EMPRESAS

DANIEL TEIXEIRA DE FARIAS

**ANÁLISE DA CAPACIDADE PREDITIVA DE MODELOS DE REGULARIZAÇÃO
APLICADO AO IPCA**

FORTALEZA

2021

DANIEL TEIXEIRA DE FARIAS

ANÁLISE DA CAPACIDADE PREDITIVA DE MODELOS DE REGULARIZAÇÃO
APLICADO AO IPCA

Dissertação de Mestrado apresentada ao curso de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de Mestre em Economia. Área de Concentração: Economia de Empresas.

Orientador: Prof. Dr. Sérgio Aquino de Souza.

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- F238a Farias, Daniel Teixeira de.
Análise da capacidade preditiva de modelos de regularização aplicado ao IPCA / Daniel Teixeira de Farias. – 2021.
38 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Mestrado Profissional em Economia de Empresas, Fortaleza, 2021.
Orientação: Prof. Dr. Sérgio Aquino de Souza.
1. Regularização. 2. Aprendizado de máquina. 3. Previsão IPCA. 4. RIDGE. 5. LASSO. I. Título.
CDD 330
-

DANIEL TEIXEIRA DE FARIAS

ANÁLISE DA CAPACIDADE PREDITIVA DE MODELOS DE REGULARIZAÇÃO
APLICADO AO IPCA

Dissertação de Mestrado apresentada ao curso de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de Mestre em Economia. Área de Concentração: Economia de Empresas.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Sérgio Aquino de Souza (Orientador)
Universidade Federal do Ceará (CAEN/UFC)

Prof. Dr. Christiano Modesto Penna
Universidade Federal do Ceará (DTE/UFC)

Prof. Dr. Rafael Barros Barbosa
Universidade Federal do Ceará (DEA/UFC)

Dedico este trabalho à minha família.

AGRADECIMENTOS

Aos meus pais e meus familiares por sempre apoiarem minhas decisões.

Ao Prof. Dr. Sérgio Aquino de Souza pelo tema proposto e pela valiosa ajuda na elaboração deste trabalho.

Aos professores integrantes do Programa de Economia Profissional que lecionaram para a turma.

Aos professores membros da banca.

Aos colegas da turma de mestrado, por terem tornado esta caminhada mais leve e prazerosa.

À Fecomércio pela parceria junto ao CAEN que possibilitou a realização deste mestrado.

A todos que tiveram alguma contribuição na minha formação.

“Todas as vitórias ocultam uma abdicação.”

(Simone de Beauvoir)

RESUMO

O presente estudo analisa o poder preditivo de curto prazo dos métodos de regularização RIDGE e LASSO aplicado ao IPCA. Com bases de dados disponíveis cada vez maiores e mais complexas, o uso de metodologia de Aprendizado de Máquina pode ser uma ótima solução em situações nas quais não se sabe previamente quais variáveis independentes devem-se adicionar no modelo de regressão. A dissertação mostra que o resultado do Erro Quadrático Médio de ambos os métodos supera significativamente o método tradicional de Mínimos Quadrados Ordinários (MQO). No período analisado, a regressão LASSO obteve inclusive um melhor resultado do que o *benchmarking* do relatório Focus, divulgado pelo Banco Central.

Palavras-chave: Regularização. Aprendizado de máquina. Previsão IPCA. RIDGE. LASSO.

ABSTRACT

This study analyzes the short-term predictive power of the Shrinkage methods RIDGE and LASSO applied to the IPCA. With increasingly larger and more complex databases available, the use of Machine Learning methodology can be a great solution in situations in which it is not previously known which independent variables should be added to the regression model. The dissertation shows that the result of the Mean Square Error of both methods significantly surpasses the traditional method of Ordinary Least Squares (OLS). In the period analyzed, the LASSO regression obtained even a better result than the benchmarking in the Focus report, released by the Central Bank.

Keywords: Shrinkage. Machine Learning. IPCA Forecast. RIDGE. LASSO.

LISTA DE FIGURAS

Figura 1 – Classificação dos métodos de Aprendizado de Máquina	17
Figura 2 – <i>Trade-off</i> viés-variância	20
Figura 3 – Retorno da função <i>ic.glmnet</i> para o método LASSO, utilizando critério BIC .	28
Figura 4 – Gráficos base de teste IPCA versus predições	29

LISTA DE GRÁFICOS

Gráfico 1 – IPCA entre Jan/2003 até Dez/2019	15
--	----

LISTA DE TABELAS

Tabela 1 – Divisão da base amostral	27
Tabela 2 – MSE da base de teste	29

LISTA DE ABREVIATURAS E SIGLAS

IPCA	Índice Nacional de Preços ao Consumidor Amplo
IBGE	Instituto Brasileiro de Geografia e Estatística
BACEN	Banco Central do Brasil
MQO	Mínimos Quadrados Ordinários
SNIPC	Sistema Nacional de Índices de Preços ao Consumidor
IBC-Br	Índice de Atividade Econômica do Banco Central
ICC	Índice de Confiança do Consumidor
ICEA	Índice de Condições Econômicas Atuais
IEF	Índice de Expectativas Futuras
CDB	Certificado de Depósito Bancário
TJLP	Taxa de Juros de Longo Prazo
CDI	Certificado de Depósito Interbancário
DLSP	Dívida Líquida do Setor Público
IGPM	Índice Geral de Preços do Mercado
IGP-DI	Índice Geral de Preços – Disponibilidade Interna
IGP 10	Índice Geral de Preços 10
SIUP	Serviços Industriais de Utilidade Pública
BNDES	Banco Nacional de Desenvolvimento Econômico e Social
TR	Taxa Referencial
NFSP	Necessidade de Financiamento do Setor Público

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Objetivos	14
1.2	Dados	14
1.3	Organização do texto	15
2	MÉTODOS EMPÍRICOS	16
2.1	IPCA – Índice de Preços ao Consumidor Amplo	16
2.2	Aprendizado de Máquina	17
2.2.1	<i>Subconjuntos de treinamento e teste</i>	18
2.2.2	<i>Regressão linear e método dos Mínimos Quadrados Ordinários</i>	20
2.2.3	<i>Avaliação da precisão do modelo</i>	22
2.3	Métodos de Regularização	23
2.3.1	<i>Regressão RIDGE</i>	23
2.3.2	<i>Regressão LASSO</i>	24
3	SIMULAÇÕES E RESULTADOS	26
3.1	Análise preliminar	26
3.2	Método e resultados	27
4	CONCLUSÃO	30
	REFERÊNCIAS	31
	APÊNDICE A – VARIÁVEIS UTILIZADAS NAS REGRESSÕES	33
	APÊNDICE B – RESULTADO DO TESTE DE ESTACIONARIDADE	36

1 INTRODUÇÃO

A inflação é um dos principais indicadores econômicos de bem-estar da sociedade, além de ser essencial no contexto de formulação de política monetária. Processos inflacionários causam destruição no poder de compras dos indivíduos, aumentam a volatilidade de preços relativos e prejudicam o planejamento dos agentes econômicos.

Devido sua criticidade econômica em tantos setores da sociedade, a elaboração de modelos preditivos de inflação possui extrema importância na tomada de decisões. Nos últimos anos, a literatura de modelos econométricos apresentou novos métodos que podem ser aplicados ao processo preditivo da inflação. Um dos motivos do desenvolvimento desses novos métodos têm ligação com o fato de que estamos diante de bases de dados cada vez maiores e mais complexas, fazendo com que os testes e a escolha dos modelos de forma otimizada sejam custosos.

A otimização no processo de modelagem se torna essencial nas situações em que o volume de dados disponível for muito extenso e/ou complexo. Isso porque, essa otimização, baseada em métricas de avaliação, poderá encontrar um modelo que pode ser classificado como melhor quando comparado aos demais. Por se tratar de abordagens orientadas aos dados, é possível utilizar ferramentas e lógicas de Aprendizado de Máquina para se ter maior eficiência no processamento das informações disponíveis, com o objetivo de encontrar melhores resultados preditivos.

No Brasil, embora o Plano Real tenha sido bem-sucedido no controle da inflação, ainda é observado uma forte persistência e instabilidade estrutural, não sendo possível comparar a variação aqui observada com o de países desenvolvidos. Por conta de países com economias emergentes apresentarem maior volatilidade na inflação, o exercício de previsão de curto prazo se torna crucial para tomadas de decisão tanto do Governo quanto do mercado, em termos de investimento.

Na literatura para estimativas de inflação no Brasil, há diferentes abordagens. Em um estudo de Arruda, Ferreira e Castelar (2011) foi utilizado modelos lineares e não lineares de séries temporais e da curva de Phillips. Os autores mostraram que, entre os mecanismos avaliados, alguns modelos não lineares e um modelo auto regressivo (AR) podem gerar erros de previsão menores do que a curva de Phillips. Barbosa, Ferreira e Silva (2020) utilizaram modelos fatoriais combinados com técnicas de aprendizado estatístico para aumentar o poder de previsão de algumas variáveis macroeconômicas, entre elas o IPCA. Garcia, Medeiros e Vasconcelos (2017) aplicaram modelos de aprendizado de máquina em dados do IPCA para

comparar sua capacidade preditiva com outras metodologias. Este último estudo mencionado possui grande influência no presente trabalho, principalmente no que diz respeito à escolha das variáveis independentes e ao método utilizado.

1.1 Objetivos

Métodos e abordagens tradicionais utilizados em econometria às vezes não são os mais eficientes para a previsão de alguns indicadores macroeconômicos, principalmente quando se tem muitas variáveis que afetam o indicador em estudo. Um tratamento que pode ser feito nestes casos é utilizar métodos de Aprendizado de Máquina que buscam melhores formas de fazer projeções, utilizando abordagem orientada a dados.

Desta forma, o principal objetivo deste trabalho consiste em testar o potencial dos modelos de regressão RIDGE e LASSO na capacidade preditiva de curto prazo do IPCA. Para avaliar essa capacidade, os resultados obtidos das simulações serão comparados com as projeções divulgadas pelo Banco Central do Brasil no relatório FOCUS.

1.2 Dados

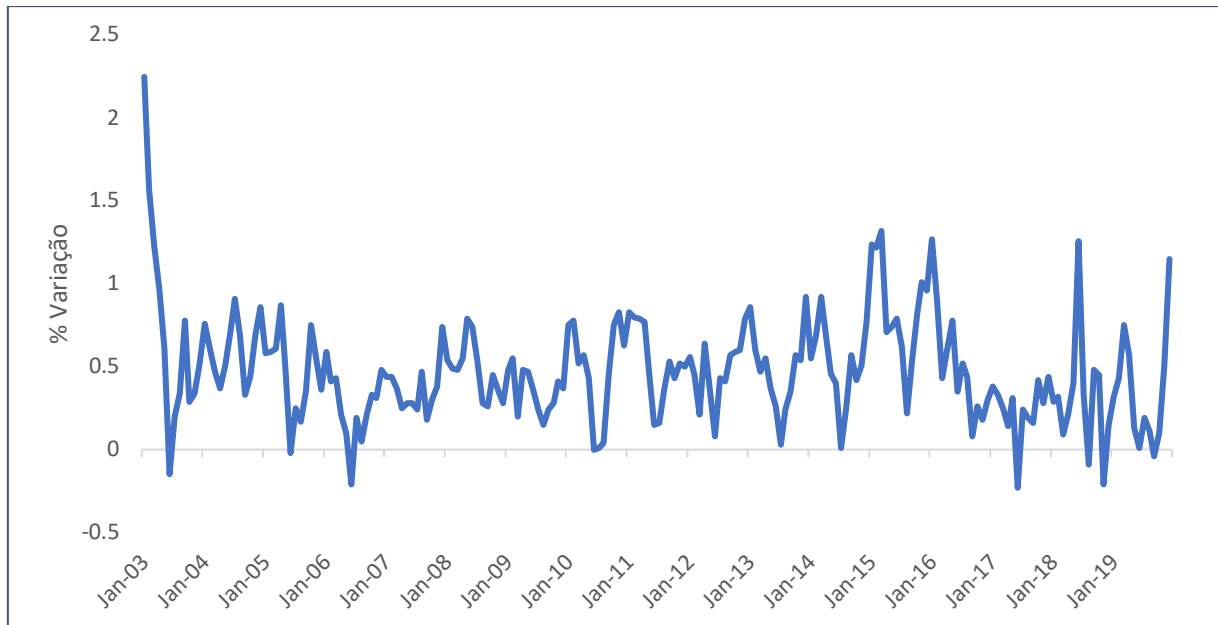
O Índice Nacional de Preços ao Consumidor Amplo (IPCA) será utilizado como indicador de inflação. O Instituto Brasileiro de Geografia e Estatística (IBGE) é o responsável pelo cálculo e divulgação dos resultados oficiais do IPCA, que é considerado o índice de inflação oficial pelo Governo Federal. Serão utilizados diversos indicadores macroeconômicos como variáveis independentes do modelo de regressão. A lista de variáveis está descrita no Apêndice A. Esses indicadores foram extraídos do Sistema Gerenciador de Séries Temporais – v2.1 (BANCO CENTRAL DO BRASIL). Foram utilizados também dados do Sistema de Expectativas do Mercado (BANCO CENTRAL DO BRASIL), que servem como referência para a geração do relatório Focus. Além disso, foi utilizado o próprio resultado do IPCA defasado em um mês.

Por se tratar de séries temporais, antes de feita as regressões, foi dada a devida tratativa, garantindo que as séries fossem estacionárias. Na seção de simulações, este procedimento será descrito com mais detalhes.

O período analisado foi entre janeiro de 2003 e dezembro de 2019 (total de 204 observações), sendo que a primeira observação foi retirada das simulações, já que foi utilizado o IPCA defasado, conforme mencionado. Nas regressões, portanto, foi utilizado o período

entre fevereiro de 2003 e dezembro de 2019. A variação percentual do IPCA neste período está mostrada no gráfico 1. Todas as simulações foram feitas utilizando a linguagem R, visto que ela oferece uma grande variedade de bibliotecas e funções já pré-definidas, possibilitando que as simulações sejam feitas de forma mais eficiente.

Gráfico 1 - IPCA entre Jan/2003 até Dez/2019



Fonte: Elaborado pelo autor a partir dos dados do IBGE.

1.3 Organização do texto

Além dessa introdução, o presente trabalho possui mais três seções. Na seção dois, será feita uma descrição teórica do IPCA e seus aspectos considerados relevantes para o desenvolvimento do trabalho. Também será abordado o conceito de aprendizado de máquina, bem como os métodos de regularização utilizados durante a simulação dos dados. Na seção seguinte, apresentam-se uma análise e descrição do banco de dados e a apresentação dos resultados obtidos do projeto. Por fim, a seção quatro será destinada a análise dos resultados e as considerações finais.

2 MÉTODOS EMPÍRICOS

Nos tópicos abaixo serão expostos os conceitos necessários para auxiliar na compreensão do presente trabalho. É importante ressaltar que, obviamente, há um vasto conteúdo que não será mencionado, uma vez que o objetivo é transmitir somente a fundamentação necessária para o entendimento da dissertação.

2.1 IPCA – Índice de Preços ao Consumidor Amplo

Conforme mencionado anteriormente, o IPCA foi utilizado nesta dissertação como índice de inflação. Ele é “utilizado pelo BACEN, desde janeiro de 1999, como parâmetro principal no monitoramento do sistema de metas de inflação no Brasil” (SITEMA, 2020, p. 34).

O indicador faz parte do Sistema Nacional de Índices de Preços ao Consumidor (SNIPC). O SNIPC foi criado em 1979, com o “[...] objetivo inicial de garantir a produção regular de medidas para captar a variação do custo de vida no Brasil” (SISTEMA, 2020, p. 8). Além do IPCA, o SNIPC produz outros índices de preços ao consumidor. Essa elaboração é feita acompanhando a variação de preços das cestas de produtos e serviços das famílias.

A definição das cestas de produtos do SNIPC é feita a partir da investigação dos hábitos de consumo das famílias, segundo a distribuição de rendimento. Esses dados de distribuição de rendimentos são utilizados, então, para definir as populações-objetivo capazes de atender a utilização proposta para cada índice (SISTEMA, 2020, p. 12).

A medição do IPCA abrange 16 áreas do Brasil: regiões metropolitanas de Belém (PA), Fortaleza (CE), Recife (PE), Salvador (BA), Belo Horizonte (MG), Rio de Janeiro (RJ), São Paulo, (SP), Curitiba (PR), Vitória (ES), Porto Alegre (RS) e as capitais Goiânia (GO), Campo Grande (MS), Rio Branco (AC), São Luís (MA), Aracaju (SE) e Brasília (DF). Visto que não é possível que a medição do índice englobasse todo o território nacional, uma vez que “[...] aumentaria a complexidade da coleta de preços, e, conseqüentemente, o montante de recursos demandados para sua realização” (SISTEMA, 2020, p. 15), as áreas acima citadas são pesquisadas por possuírem um maior volume de transações monetárias e uma melhor estrutura formal de comercialização, o que propicia um acompanhamento constante da evolução de preços praticados. Dessa forma, a pesquisa se restringe a famílias residentes na área urbana dos Estados. Normalmente, os preços são coletados do 1 ao 30 dia do mês e comparados aos observados no mesmo período do mês anterior.

2.2 Aprendizado de Máquina

O Aprendizado de Máquina é uma área da Inteligência Artificial que automatiza a criação de modelos a partir de análise de dados. Ou seja, não há necessidade de se definir explicitamente um algoritmo com todas as instruções detalhadas, basta fornecer dados para que o computador realize as tarefas e tome as decisões.

Embora não haja consenso na literatura quanto à classificação dos métodos de Aprendizado de Máquina, o mais comum é categorizar entre aprendizagem supervisionada, não supervisionada ou por reforço. Neste, há uma interação entre o agente (que faz as ações) e o ambiente (que retorna recompensas) a fim de maximizar as recompensas totais. No aprendizado não supervisionado é abordado problemas onde não há uma variável-objetivo específica a ser alcançada, buscando-se identificação de grupos ou padrões a partir dos dados de entrada. Já no aprendizado supervisionado, busca-se uma variável-objetivo, com o modelo utilizando os valores passados da variável-objetivo para aprender quais devem ser seus resultados de saída. É no aprendizado supervisionado onde se encontram as regressões, como pode ser visto na figura 1, e, portanto, os métodos utilizados neste trabalho se inserem nesta classificação.

Figura 1 – Classificação dos métodos de Aprendizado de Máquina



Fonte: Gao et al. (2020), com adaptações.

É importante destacar que quando se menciona Aprendizado de Máquina e métodos orientados a dados, se refere a uma forma de raciocínio que usa diferentes métodos para o alcance dos objetivos. Não se restringe, necessariamente, a um conjunto limitado de metodologias estatísticas e computacionais. Os objetivos principais são o desempenho preditivo dos modelos e a automatização do processo de modelagem das bases de dados.

As análises no campo da economia utilizando a metodologia de Aprendizado de Máquina têm evoluído significativamente nos últimos anos. Segundo Varian (2014, p. 1), “esses métodos têm muito a oferecer e devem ser amplamente conhecidos e utilizados por economistas”. Bajari et al. (2015) destaca que embora alguns modelos possam ser desconhecidos para alguns economistas, eles são logicamente simples, já que são baseados em métodos considerados mais tradicionais. Assim, entre as justificativas para o aumento de interesse, destaca-se:

- (i) Elevado grau de automação do processo.

A automatização nos processos de modelagem, estimação, teste e previsão traz muitas vantagens em áreas que tratam de previsão.

- (i) Atua diretamente em problemas de *big data*.

Situações em que há uma grande base de dados com um número elevado de preditores relativo às observações disponíveis estão cada vez mais presentes em vários campos, inclusive na área de economia. No caso de muitos preditores, existe uma possibilidade considerável de haver multicolinearidade elevada entre eles.

2.2.1 Subconjuntos de treinamento e teste

Para medir o desempenho do modelo, é necessário separar a base de dados em dois grupos: treinamento e teste. O primeiro é o responsável pelo aprendizado do modelo (definição dos coeficientes), enquanto que o segundo, como o próprio nome indica, serve para que se realizem os testes do modelo.

Normalmente, estamos interessados em saber o desempenho do algoritmo de aprendizado de máquina em dados que ele não viu antes, uma vez que isso determina o quão bem ele funcionará quando implementado no mundo real. Portanto, avaliamos essas medidas de desempenho usando um conjunto de dados de teste separado dos dados usados para treinar o sistema de aprendizagem da máquina (GOODFELLOW et al., 2016, p. 102).

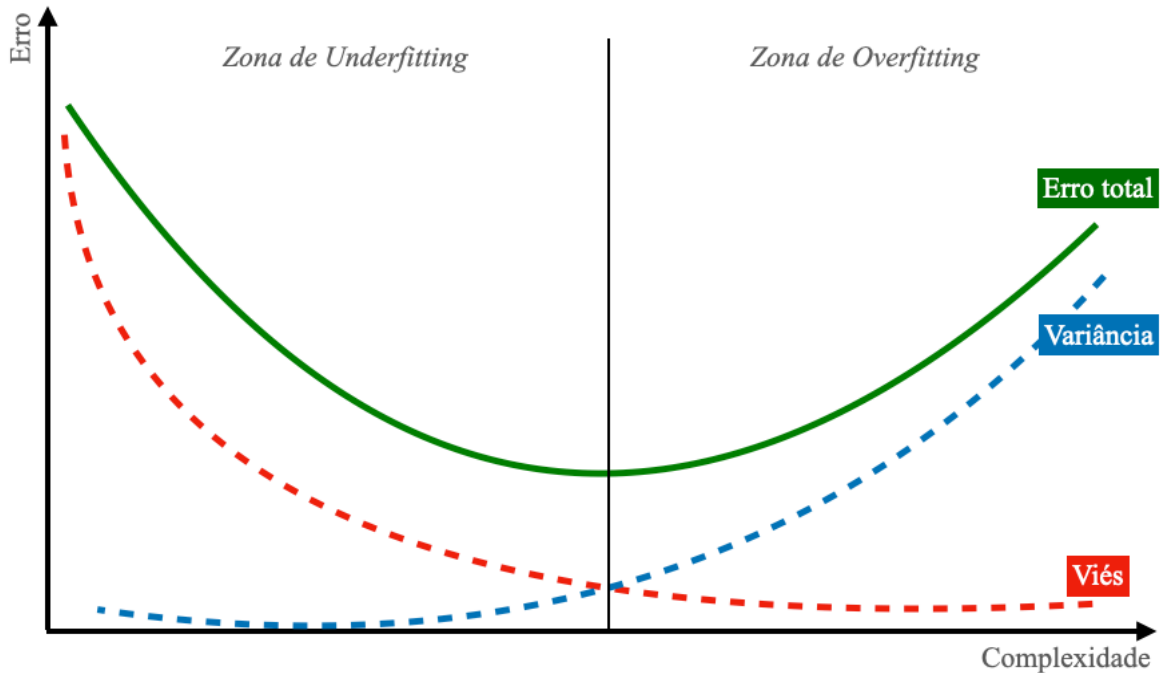
Além disso, quando se trata de séries temporais, a divisão não pode ser feita de forma aleatória, devendo seguir o critério temporal da série. Portanto, os dados da base de treino devem vir antes dos dados da base de teste.

Caso o procedimento de divisão não fosse realizado e toda a base fosse utilizada, o modelo criado poderia apresentar alta variância, se ajustando excessivamente aos dados, inclusive aos erros e fatores aleatórios. Modelos com essas características têm problema de *overfitting*, com capacidade de generalização ruim, já que houve ajuste exagerado aos dados de treino.

O oposto ocorre quando o modelo tem pouca aderência à base de treino, causando *underfitting*. Neste caso, o modelo obtido é simplista, não registrando corretamente a complexidade da base de dados, possuindo assim um alto viés.

James et al. (2013) apresentam uma minuciosa discussão acerca do *trade-off* viés-variância. Para o presente trabalho, é suficiente ter em mente os conceitos expostos, além de que o equilíbrio entre viés e variância é um dos principais desafios a serem enfrentados na elaboração de modelos de predição. Ambas podem ser vistas como métricas de erro do modelo e um bom balanceamento de seus valores contribuem para alcançar uma alta acurácia preditiva. A figura 2 ilustra esse entendimento. Conforme o aumento da complexidade do modelo, o viés diminui ao preço de um aumento na variância, causando o *overfitting*, que pode ocasionar um resultado ruim na predição na base de teste. Por outro lado, um modelo com baixa complexidade produz erros altos na base de treinamento, possuindo um alto viés. A linha preta vertical traçada no ponto mais baixo da curva de “Erro total” é a situação desejada: complexidade média do modelo, com boa capacidade de predição.

Figura 2 – Trade-off viés-variância



Fonte: Elaborada pelo autor a partir de Goodfellow et al. (2016, p. 128).

Além dos erros de predição causados por viés e/ou variância, há também o erro irreduzível, que basicamente se refere à *outliers*. Pode ocorrer na elaboração do modelo que alguma variável relevante não tenha sido considerada ou que nem todas as informações estejam disponíveis, fazendo com que as predições geradas sejam incorretas. Os modelos de regressão de Aprendizado de Máquina empenham-se em reduzir o erro da parcela redutível. Para visualizar essa separação matematicamente, conforme explica James et al. (2013), seja “ $E(Y - \hat{Y})^2$ a média, ou valor esperado, da diferença quadrática entre o valor previsto e o valor real de Y”. Então:

$$E(Y - \hat{Y})^2 = E[f(X) + \epsilon - \hat{f}(X)]^2 = [f(X) - \hat{f}(X)]^2 + Var(\epsilon) \quad (2.1)$$

Onde $\hat{Y} = \hat{f}(X)$ é a estimativa de Y. Assim, o termo $[f(X) - \hat{f}(X)]^2$ é a parcela redutível do erro e $Var(\epsilon)$ é o erro irreduzível.

2.2.2 Regressão linear e método dos Mínimos Quadrados Ordinários

O objetivo de uma regressão linear é analisar a existência de uma relação funcional entre uma variável dependente Y com um conjunto de variáveis independentes $X = (X_1, X_2, \dots, X_p)$. Matematicamente, tem-se:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon \quad (2.2)$$

no qual $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ é o vetor de parâmetros da regressão que são desconhecidos e previsão ser estimados e ϵ é o erro.

Nos modelos preditivos, ao se encontrar os parâmetros, é necessário que se faça um teste utilizando a estatística F para saber se eles são significantes ou não.

Uma das abordagens mais tradicionais da literatura para ajustes de modelos de regressão linear é utilizando a técnica dos Mínimos Quadrados Ordinários (MQO), que propõe uma solução na qual minimiza a soma dos quadrados das diferenças entre o valor estimado e os dados observados.

Considerando que $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \dots + \hat{\beta}_p x_p$ é a previsão para y com base no i -ésimo valor de x . Então, $e_i = y_i - \hat{y}_i$ representa o i -ésimo resíduo, que é a diferença entre o i -ésimo valor da resposta observada e o i -ésimo valor previsto no modelo linear. O procedimento que determina os parâmetros da regressão é encontrado usando os valores que minimizam a soma dos quadrados dos resíduos (*RSS - Residual Sum of Squares*, em inglês) que é dado por:

$$RSS = e_1^2 + e_2^2 + \dots + e_n^2$$

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (2.3)$$

Segundo o teorema de Gauss-Markov, dentre os estimadores lineares não viesados, o MQO é o que apresenta a menor variância. Embora possua uma estrutura simples, o MQO possui desempenho muito bom quando comparado a modelos mais sofisticados. Entretanto, conforme indicam Kuhn e Johnson (2013), o MQO demonstra quatro limitações nos quais os estimadores são instáveis ou não são os mais adequados:

- Situação em que há elevada multicolinearidade nos dados;
- A solução do modelo é uma aproximação linear que em muitos casos não representa adequadamente o espaço estudado;
- Estimadores são muito sensíveis a *outliers*;
- Não pode ser usado em situações em que o número de parâmetros for superior ao tamanho da amostra.

Por esses motivos citados, métodos alternativos foram desenvolvidos para serem aplicados nos casos em que os modelos de regressão tradicionais não conseguem explicar (ou

não podem ser aplicados) a relação entre as variáveis explicativa e explicada. Além disso, o MQO tem como limitação a seleção de variáveis preditoras que de fato são importantes no modelo. Isso porque, o MQO não possui um critério de seleção de um subconjunto de preditores dentro de um conjunto maior. Isso afeta diretamente a interpretabilidade do modelo.

Muitas vezes acontece que algumas ou muitas das variáveis usadas em um modelo de regressão múltipla, na verdade, não estão associadas à resposta. Incluir essas variáveis irrelevantes leva a uma complexidade desnecessária no modelo resultante. Removendo essas variáveis, ou seja, definindo as estimativas dos coeficientes correspondentes para zero, podemos obter um modelo que é mais facilmente interpretado (JAMES et al., 2013, p. 204).

Para esses casos, a literatura de Aprendizado de Máquina sugere métodos em que se automatizam o processo de seleção. Dentre esses métodos destaca-se o de regressões regularizadas.

2.2.3 Avaliação da precisão do modelo

Para avaliar o desempenho de um modelo de regressão, há várias métricas que podem ser utilizadas. A mais empregada e indicada pela literatura é a do Erro Quadrático Médio (*MSE – Mean Square Error*, em inglês), que é calculado por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{(y_1 - \hat{y}_1)^2 + \dots + (y_n - \hat{y}_n)^2}{n} \quad (2.4)$$

Sendo n o número de observações, y_i é o i -ésimo valor de Y e \hat{y}_i é o i -ésimo valor de \hat{Y} . O MSE será baixo se as predições forem próximas aos valores atuais e aumentará o seu valor conforme as respostas previstas e atuais se distanciar. Como o desempenho do modelo é medido pela base de teste, tem-se que:

$$MSE_{teste} = \frac{1}{n} \sum_{i=1}^n (y_i^{(teste)} - \hat{y}_i^{(teste)})^2 \quad (2.5)$$

O MSE pode ser calculado tanto na base de treinamento quanto na base de teste. Em termos de seleção de modelos, o que realmente importa é o resultado na base de teste. No caso em que a base de teste é disponível, basta que a fórmula seja aplicada para ter acesso ao resultado. Entretanto, pode ocorrer situações em que apenas a base de treinamento é acessível. Neste caso, embora possa parecer uma abordagem sensata selecionar o método que minimize

o MSE da base de treinamento, isto não garante que o MSE na base de teste será o menor.

O problema de *overfitting* também pode ser constatado observando os resultados do MSE. Quando um método resulta em um pequeno MSE da base de treinamento e um grande MSE da base de teste, pode-se concluir que o modelo se ajustou excessivamente aos dados do treino. Conforme pontua James et al. (2013, p. 32), “quando os dados de treinamento são ajustados demasiadamente, o MSE de teste será muito grande porque os supostos padrões que o método encontrou nos dados de treinamento simplesmente não existem nos dados de teste”.

2.3 Métodos de Regularização

No método de MQO, o fato do número de parâmetros se aproximar ao tamanho da amostra faz com que a variância das estimativas do modelo seja elevada. Há ainda casos em que o número de parâmetros é superior ao tamanho da amostra. Neste caso, existem infinitas soluções e o MQO não pode ser aplicado. A sugestão para esses casos é usar métodos de *Shrinkage* (também conhecidos como regularização ou encolhimento). Segundo Hastie et al. (2009), estes métodos que aplicam penalidades nos coeficientes dos estimadores são mais contínuos e não apresentam elevada variabilidade. A penalidade é aplicada para limitar a variância dos estimadores ao custo de um viés nos parâmetros estimados. Segundo Kuhn e Johnson (2013), acrescentar um pequeno viés nos estimadores pode reduzir consideravelmente a variância do modelo. Goodfellow et al. (2016, p. 117) conceitua regularização como “[...] qualquer modificação que fazemos em um algoritmo de aprendizagem que se destina a reduzir seu erro”.

Esses métodos são denominados viesados porque tendem a reduzir os estimadores de mínimos quadrados em direção a zero (daí o nome encolhimento). Podem também realizar a seleção de variáveis, já que, dependendo do tipo de contração realizada, alguns dos coeficientes podem ser estimados em exatamente zero. Os dois métodos de regularização mais conhecidos são as regressões RIDGE e LASSO.

2.3.1 Regressão RIDGE

Matematicamente, a regressão RIDGE é muito semelhante ao método MQO, com a adição de um termo de penalização. As estimativas dos coeficientes são os valores que minimizam:

$$\hat{\beta}^{ridge} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2 \quad (2.6)$$

O segundo termo da equação $\lambda \sum_{j=1}^p \beta_j^2$ é chamado de *shrinkage penalty*. O coeficiente λ , chamado de parâmetro de regularização, controla o impacto relativo dos termos nas estimativas do coeficiente de regressão e é determinado separadamente. Quando $\lambda = 0$, o termo de penalização não afeta a regressão e o resultado é o mesmo obtido no método dos MQO. Ou seja, o estimador de MQO pode ser visto como um caso particular do estimador via RIDGE (e também via LASSO). Conforme $\lambda \rightarrow \infty$, o impacto do termo de penalização aumenta, fazendo com que o coeficiente estimado da regressão RIDGE se aproxime de zero. Percebe-se assim que a regressão RIDGE possui uma característica de suavizar os atributos que têm menor contribuição na predição do modelo. É importante notar que o segundo termo da equação só será exatamente igual a zero no caso de $\lambda = \infty$.

A regressão RIDGE produzirá um conjunto diferente de estimativas de coeficientes β_{λ}^{ridge} para cada valor de λ , sendo este parâmetro responsável pela intensidade da restrição imposta para cada coeficiente. Outro ponto importante é o fato de que o *shrinkage penalty* não é aplicado no intercepto β_0 , apenas nos coeficientes β_1, \dots, β_p .

Conforme James et al. (2013, p. 217), “vantagem da regressão RIDGE sobre os mínimos quadrados está enraizada no *trade-off* viés-variância. À medida que λ aumenta, a flexibilidade do ajuste de regressão RIDGE diminui, levando à diminuição da variância, mas aumentando o viés”. Uma das principais críticas à regressão RIDGE é que ela não exclui nenhuma variável preditora, já que não tem o poder de zerar os coeficientes. Portanto, no resultado estará incluído todos os coeficientes ou nenhum deles.

2.3.2 Regressão LASSO

A regressão LASSO é bastante similar à regressão RIDGE, sendo que uma das diferenças é que é possível encontrar coeficientes que sejam iguais a zero. Matematicamente, as estimativas dos coeficientes são os valores que minimizam:

$$\hat{\beta}^{lasso} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = RSS + \lambda \sum_{j=1}^p |\beta_j| \quad (2.7)$$

Comparando as fórmulas RIDGE e LASSO, é fácil notar que a única diferença está no *shrinkage penalty*. Enquanto que na RIDGE é utilizado o termo β_j^2 , na LASSO o termo é substituído por $|\beta_j|$. Dependendo do valor do parâmetro de regularização λ , as estimativas de alguns dos coeficientes podem ser iguais a zero. Entende-se desta forma que a regressão LASSO pode realizar seleção de variáveis. Por conta disso, geralmente tem-se como resultado que os modelos gerados por LASSO são mais fáceis de interpretar que os modelos gerados por RIDGE. Outra observação importante é que, assim como ocorre na regressão RIDGE, na LASSO também se obtém como resultado um conjunto diferente de estimativas de coeficientes para cada valor de λ .

3 SIMULAÇÕES E RESULTADOS

3.1 Análise preliminar

As variáveis independentes escolhidas para simulação do modelo se encontram nas seguintes categorias: atividade econômica, setor externo, mercado financeiro e de capitais, finanças públicas, indicadores monetários, mercado de trabalho e expectativa de mercado. Conforme mencionado anteriormente, o período analisado foi entre fevereiro de 2003 e dezembro de 2019, com cada mês sendo uma observação, fazendo um total de 203 itens.

Visto que os dados de estudo desta dissertação são séries temporais, foi necessário primeiramente verificar sua estacionariedade antes de fazer qualquer modelagem preditiva. Uma série temporal estacionária possui média, variância e autocorrelação constantes ao longo do tempo. Cada uma dessas propriedades, quando não observadas, podem interferir nos resultados de regressão.

Caso a média não seja constante, os modelos de regressão podem não ser eficientes, já que os valores previstos podem se distanciar significativamente da média. Em caso de variância não constante, os erros de previsão podem ser maiores em determinados períodos, uma vez que se espera que ela se mantenha inconstante ao longo do tempo. Por fim, caso não haja autocorrelação entre o período atual com períodos anteriores, estaríamos diante de uma série com valores aleatórios e imprevisíveis.

Assim, para verificar a existência de estacionariedade em cada uma das variáveis independentes utilizadas nas regressões foi feito o teste de Dickey-Fuller Aumentado (*ADF – Augmented Dickey-Fuller*, em inglês). O teste ADF investiga a existência ou não de raiz unitária. Resumidamente, a hipótese testada e sua relação com estacionariedade pode ser descrita como:

- H0: existe raiz unitária (não é estacionária);
- H1: não existe raiz unitária (é estacionária).

Dessa forma, considerando um nível de confiança de 95%, deseja-se que o p-valor encontrado seja menor do que 0,05. Os resultados obtidos estão no Apêndice B. Nos casos em que o teste constatou a não existência de estacionariedade, foi utilizada a metodologia de diferenciação, conforme amplamente indicado pela literatura. Ao fazer a primeira diferenciação, o efeito acumulado da série é retirado, ficando a diferença entre o valor no período t com o valor no período $t-1$. Nas variáveis em que foram constatadas não

estacionariedade, após feita a primeira diferença, foi feito novamente o teste ADF para cada uma delas e os resultados também estão no Apêndice B.

Por fim, antes de iniciar a simulação dos modelos, a base de dados foi dividida em conjuntos de treinamento e teste. A literatura indica que entre 70% e 80% da amostra total seja agrupada na base de treinamento e o restante (20% a 30%) fique na base de teste. Considerando que o período em estudo foi entre fevereiro de 2003 e dezembro de 2019 (total de 203 observações), a divisão foi feita conforme a tabela 1:

Tabela 1 - Divisão da base amostral

	Número de observações	Referência
Base de treinamento	155	Fev/2003 a Dez/2015
Base de teste	48	Jan/2016 a Dez/2019

Fonte: Elaborada pelo autor.

3.2 Método e resultados

Conforme mencionado, a simulação dos modelos foi feita via linguagem R. Dois pacotes foram utilizados:

- *glmnet*: criado por Friedman et al. (2021), é um pacote que ajusta modelos lineares e outros similares via máxima verossimilhança;
- *HDeconometrics*: criado por Vasconcelos, G. (2016), utiliza o pacote *glmnet* para a estimação via RIDGE e LASSO e seleciona o melhor modelo utilizando o critério de informação de acordo com a escolha do usuário.

No que diz respeito aos métodos de regularização RIDGE e LASSO, uma das principais tarefas é a de encontrar um valor ótimo para λ . O método mais comum para selecionar este parâmetro é utilizando validação cruzada. Entretanto, diferentes estudos feitos por Wang et al. (2007), Zou et al. (2007) e Zhang et al. (2010) usaram Critério de Informação Bayesiano (*BIC - Bayesian Information Criteria*, em inglês) para mostrar que a escolha de λ via BIC é consistente. Conforme afirma Medeiros e Mendes (2016, p. 5), a utilização do BIC “oferece resultados superiores tanto em termos de precisão quanto em tempo de computação, do que métodos de validação cruzada”.

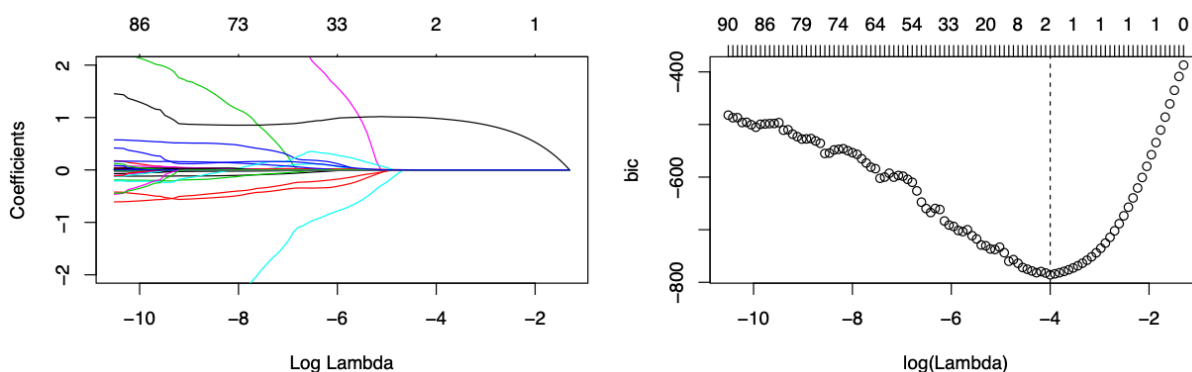
Desconsiderando todo o tratamento da base de dados na análise preliminar, as duas principais funções utilizadas para simular as regressões e se obter os respectivos

coeficientes foram:

- (i) *ic.glmnet*: faz o ajuste do modelo, escolhendo o melhor λ a partir dos argumentos:
 - a) Matriz X que deve conter as variáveis independentes;
 - b) Vetor Y que deve conter o vetor da variável objetivo;
 - c) Parâmetro α que determina o método utilizado ($\alpha = 0$ para RIDGE; $\alpha = 1$ para LASSO);
 - d) Parâmetro que determina o critério a ser usado (no caso deste estudo, BIC).
- (ii) *predict*: realiza as previsões dos modelos ajustados por *ic.glmnet* e tem os seguintes argumentos:
 - a) O modelo que se deseja ajustar;
 - b) Nova matriz de variáveis independentes, ou seja, a base de teste.

A figura 3 ilustra o funcionamento e mostra o retorno da função *ic.glmnet*, para o caso do método LASSO. O gráfico da esquerda exibe as variáveis independentes indo para zero conforme a penalização aumenta. Já o gráfico da direita representa o BIC selecionando o melhor modelo com base no valor de λ .

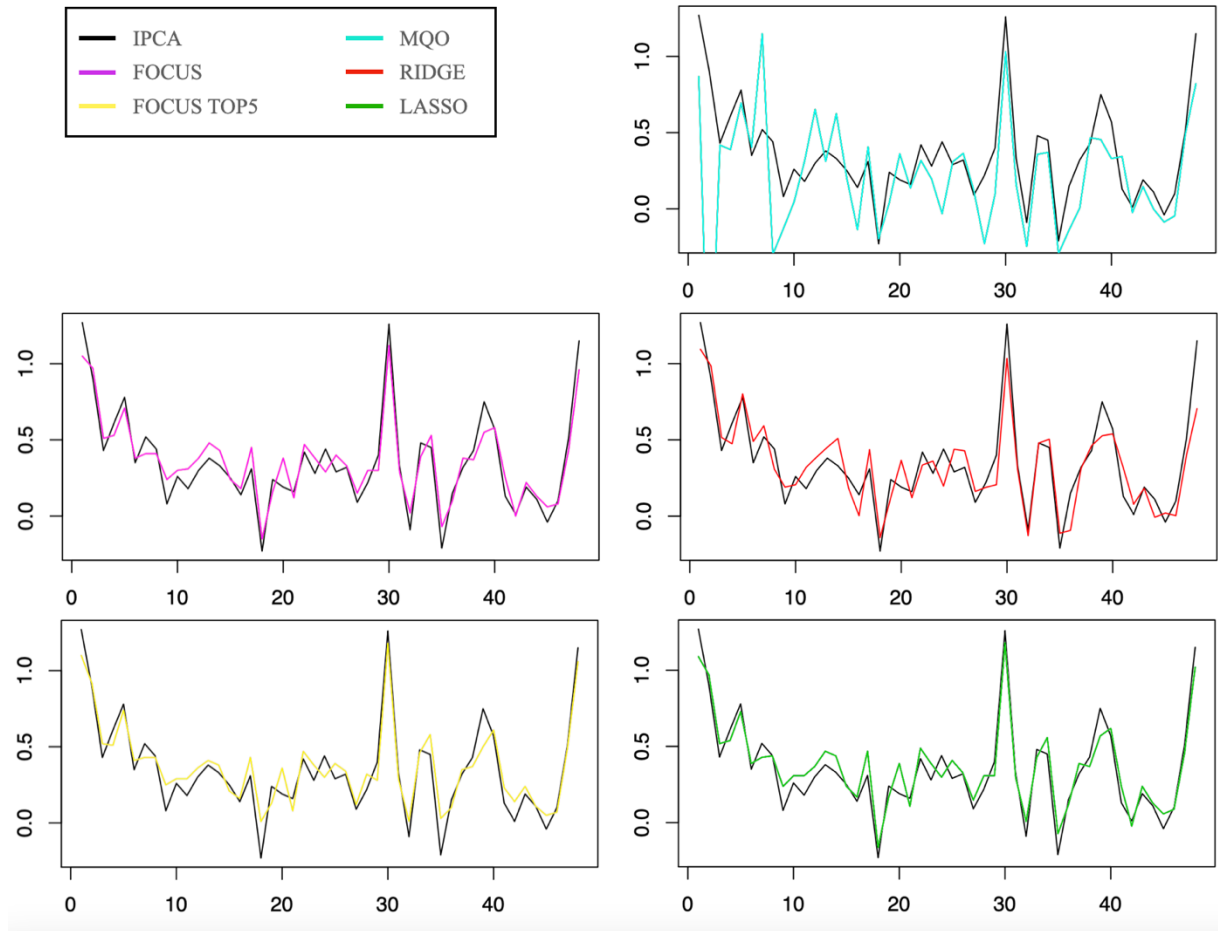
Figura 3 – Retorno da função *ic.glmnet* para o método LASSO, utilizando critério BIC



Fonte: Elaborada pelo autor.

O resultado da função *predict* pode ser visto na figura 4, que mostra o comportamento das 48 observações da base de teste do IPCA comparado com as previsões Focus e também com os métodos de regressão utilizados nesta dissertação. Embora a análise visual não seja concludente no que diz respeito à avaliação da precisão do modelo, é importante pontuar a disparidade entre o MQO com todos os outros resultados.

Figura 4 – Gráficos base de teste IPCA versus predições



Fonte: Elaborada pelo autor.

A tabela 2 mostra os resultados de MSE das 48 observações da base de teste. A previsão feita via MQO foi a que apresentou maior MSE dentre os métodos de regressão. Embora o resultado do RIDGE tenha sido ainda maior do que o observado no FOCUS, é importante notar o ganho quando comparado ao MQO, mesmo mantendo todos os coeficientes, já que diferentemente do LASSO nenhuma das estimativas é igual a zero. Este ganho frente ao método tradicional de MQO pode ser atribuído ao *trade-off* viés-variância. Por fim, o resultado de MSE da regressão LASSO foi o menor obtido, mostrando ganhos significativos também quando comparado ao observado no FOCUS.

Tabela 2 - MSE da base de teste

MQO	RIDGE	FOCUS TOP5	FOCUS	LASSO
0.175531837	0.018506850	0.010620833	0.010006250	0.009075505

Fonte: Elaborada pelo autor.

4 CONCLUSÃO

Com o volume de dados disponível crescendo exponencialmente e o avanço de áreas de estudo como o de aprendizado de máquina, os métodos de regularização são uma ótima alternativa num contexto em que não seja possível se fazer uma análise prévia de quais variáveis independentes devem ser colocadas num modelo de regressão. Neste caso, métodos tradicionais como o MQO podem não retornar a melhor resposta, uma vez que algumas dessas variáveis independentes podem não estar associadas com a variável objetivo.

Os resultados de erro quadrático médio das regressões RIDGE e LASSO foram notavelmente melhores do que o obtido no método tradicional de MQO na amostra do IPCA estudado, mostrando que o *trade-off* viés-variância pode trazer ganhos preditivos quando utilizado de forma equilibrada. Além disso, a previsão utilizando LASSO obteve resultado ainda melhor do que a comparação com o boletim FOCUS.

Importante notar que esses métodos de regularização são relativamente recentes e que novas adaptações têm sido propostas por diferentes autores. Como exemplos, pode ser citado o adaLASSO (*Adaptative LASSO*), WLadaLASSO (*Weighted Lag Adaptative LASSO*) e Post-LASSO. Há também o *Elastic Net*, que pode ser visto como uma combinação entre o RIDGE e o LASSO. Desta forma, como sugestão de trabalhos futuros, poderia ser incluído esses modelos para comparar a previsão do IPCA com os resultados obtidos. Outra proposta seria estender o período de previsão e observar os retornos de cada uma dessas metodologias.

REFERÊNCIAS

- ARRUDA, E.; FERREIRA, R.; CASTELAR, I. Modelos lineares e não lineares da curva de Phillips para previsão da taxa de inflação no Brasil. **Revista Brasileira de Economia**, [s. l.], v. 65, n. 3, p. 237-252, 2011.
- BAJARI, P.; NEKIPOLEV, D.; RYAN, S.; YANG, M. Machine learning methods for demand estimation. **American Economic Review**, [s. l.], v. 105, n. 5, p. 481-485, 2015.
- BANCO CENTRAL DO BRASIL. **SGS - Sistema Gerenciador de Séries Temporais - v2.1**. Disponível em: <<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocprepararTela>>. Acesso em: 01 abril 2021.
- BANCO CENTRAL DO BRASIL. **Sistema de expectativas de mercado**. Disponível em: <<https://www3.bcb.gov.br/expectativas/publico/consulta/serieestatisticas>>. Acesso em: 01 abril 2021.
- BARBOSA, R. B.; FERREIRA, R.; SILVA, T. M. Previsão de variáveis macroeconômicas brasileiras usando modelos de séries temporais de alta dimensão. **Estud. Econ.**, São Paulo, v. 50, n. 1, p. 67-98, 2020.
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R.; NARASIMHAN, B.; TAY, K.; SIMON, N. **Package ‘glmnet’**. CRAN R Repository, 2021. Disponível em: <<https://cran.r-project.org/web/packages/glmnet/glmnet.pdf>>. Acesso em: 01 abril 2021.
- GAO, K.; MEI, G.; PICCIALLI, F.; CUOMO, S.; TU, J.; HUO, Z. Julia language in Machine learning: algorithms, applications, and open issues. **Computer Science Review**, [s. l.], v. 37, p. 100254, 2020.
- GARCIA, M.; MEDEIROS, M.; VASCONCELOS, G. Real-time inflation forecasting with high-dimensional models: The case of Brazil. **International Journal of Forecasting**, [s. l.], v. 33, p. 679-693, 2017.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>. Acesso em: 01 abril 2021.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference and prediction**. Springer Series in Statistics, 2009.
- HASTIE, T.; QIAN, J.; TAY, K. **An introduction to glmnet**. CRAN R Repository, 2021. Disponível em: <<https://cran.r-project.org/web/packages/glmnet/vignettes/glmnet.pdf>>. Acesso em: 01 abril 2021.
- IBGE. **Sistema nacional de índice de preços ao consumidor**. Série Relatórios Metodológicos, 8 ed, Rio de Janeiro, 2020.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning with applications in R**. Springer Texts in Statistics, 2013.

KUHN, M.; JOHNSON, K. **Applied predictive modeling**. Springer, 2013.

MEDEIROS, M.; VASCONCELOS, G.; FREITAS, E. Forecasting Brazilian inflation with high-dimensional models. **Brazilian Review of Econometrics**, [s. l.], v. 36, p. 68–100, 2016.

MEDEIROS, M.; VASCONCELOS, G. Forecasting macroeconomic variables in data-rich environments. **Economics Letters**, [s. l.], v. 138, p. 50–52, 2016.

MEDEIROS, M. C.; MENDES, E. F. **L1 - Regularization of high-dimensional time-series models with flexible innovations**, 2015.

UNIVERSIDADE FEDERAL DO CEARÁ. Biblioteca Universitária. **Modelo de trabalhos acadêmicos da Universidade Federal do Ceará**. Fortaleza, 2018.

VARIAN, H. R. Big data: new tricks for econometrics. **Journal of Economic Perspectives**, [s. l.], v. 28, n. 2, p. 3-28, 2014.

WANG, H.; LI, G.; TSAI, C. Regression coefficient and autoregressive order shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, [s. l.], v. 69, p. 63–78, 2007.

ZHANG, Y.; LI, R.; TSAI, C. Regularization parameter selections via generalized information criterion. **Journal of the American Statistical Association**, [s. l.], v. 105, p. 312–323, 2010.

ZOU, H.; HASTIE, T.; TIBSHIRANI, R. On the degrees of freedom of the lasso. **The Annals of Statistics**, [s. l.], v. 35, p. 2173–2192, 2007.

APÊNDICE A – VARIÁVEIS UTILIZADAS NAS REGRESSÕES

Variável	Categoria	Fonte
IBC-Br	Atividade econômica	BACEN
Indicadores de produção industrial	Atividade econômica	IBGE
Energia elétrica comercial	Atividade econômica	Eletrobrás
Energia elétrica residencial	Atividade econômica	Eletrobrás
Energia elétrica outros	Atividade econômica	Eletrobrás
Índice de vendas combustíveis	Atividade econômica	IBGE
Índice de vendas hipermercado	Atividade econômica	IBGE
Índice de vendas vestuário	Atividade econômica	IBGE
Índice de vendas eletrodomésticos	Atividade econômica	IBGE
Índice de vendas automóveis	Atividade econômica	IBGE
Índice de vendas supermercado	Atividade econômica	IBGE
ICC	Atividade econômica	Fecomércio
ICEA	Atividade econômica	Fecomércio
IEF	Atividade econômica	Fecomércio
Exportação de bens	Setor externo	BACEN
Importação de bens	Setor externo	BACEN
Taxa CDB	Mercado financeiro e de capitais	BACEN
Taxa CDB pessoa física	Mercado financeiro e de capitais	BACEN
Taxa CDB pessoa jurídica	Mercado financeiro e de capitais	BACEN
Taxa CDB investidores institucionais	Mercado financeiro e de capitais	BACEN
Taxa CDB pós-fixado	Mercado financeiro e de capitais	BACEN
TJLP	Mercado financeiro e de capitais	BACEN
CDI mensal	Mercado financeiro e de capitais	BACEN
DLSP Governo	Finanças públicas	BACEN
DLSP BACEN	Finanças públicas	BACEN
DLSP Estados	Finanças públicas	BACEN
DLSP Municípios	Finanças públicas	BACEN
DLSP Empresas federais	Finanças públicas	BACEN
DLSP Empresas estatais	Finanças públicas	BACEN
DLSP Empresas municipais	Finanças públicas	BACEN
IGPM	Atividade econômica	IBGE
IGPDI	Atividade econômica	IBGE
IGP10	Atividade econômica	IBGE
IPCA15	Atividade econômica	IBGE
Agregado monetário M1	Indicadores monetários	BACEN
Agregado monetário M3	Indicadores monetários	BACEN
Agregado monetário M4	Indicadores monetários	BACEN

Base monetária	Indicadores monetários	BACEN
Empregos formais	Mercado de trabalho	ME
Empregos formais agricultura	Mercado de trabalho	ME
Empregos formais indústria	Mercado de trabalho	ME
Empregos SIUP	Mercado de trabalho	ME
Empregos formais construção	Mercado de trabalho	ME
Empregos formais comércio	Mercado de trabalho	ME
Vendas industriais	Atividade econômica	CNI
Indicador indústria da transformação	Atividade econômica	IBGE
Horas trabalhadas da indústria	Atividade econômica	CNI
Utilização capacidade instalada	Atividade econômica	FGV
Taxa de câmbio US\$ compra	Setor externo	BACEN
Taxa de câmbio US\$ venda	Setor externo	BACEN
Índice Ibovespa	Mercado financeiro e de capitais	BM&FBOVESPA
Operações compromissadas SELIC	Indicadores monetários	BACEN
Títulos privados	Indicadores monetários	BACEN
Quotas de fundos monetários	Indicadores monetários	BACEN
Títulos federais no poder público	Indicadores monetários	BACEN
Desembolso BNDES	Atividade econômica	BNDES
Desembolso BNDES indústria	Atividade econômica	BNDES
Desembolso BNDES comércio	Atividade econômica	BNDES
Desembolso BNDES agricultura	Atividade econômica	BNDES
Desembolso BNDES extração	Atividade econômica	BNDES
Taxa SELIC mês	Mercado financeiro e de capitais	BACEN
Dívida mobiliária Tesouro	Finanças públicas	BACEN
Dívida mobiliária custódia Tesouro	Finanças públicas	BACEN
Dívida mobiliária custódia BACEN	Finanças públicas	BACEN
Dívida mobiliária saldo Tesouro	Finanças públicas	BACEN
Dívida mobiliária IGPDI carteira	Finanças públicas	BACEN
Dívida mobiliária IGPM carteira	Finanças públicas	BACEN
Dívida mobiliária TR	Finanças públicas	BACEN
Dívida mobiliária câmbio carteira	Finanças públicas	BACEN
Dívida mobiliária IGPDI custódia	Finanças públicas	BACEN
Dívida mobiliária IGPM custódia	Finanças públicas	BACEN
Dívida mobiliária pré-fixada	Finanças públicas	BACEN
Dívida mobiliária câmbio custódia	Finanças públicas	BACEN
NFSP Governo/BACEN	Finanças públicas	BACEN
NFSP Governo	Finanças públicas	BACEN
NFSP BACEN	Finanças públicas	BACEN
NFSP estado	Finanças públicas	BACEN
NFSP município	Finanças públicas	BACEN

NFSP empresas estatais	Finanças públicas	BACEN
NFSP empresas federais	Finanças públicas	BACEN
NFSP empresas estaduais	Finanças públicas	BACEN
NFSP empresas municipais	Finanças públicas	BACEN
NFSP fluxo Governo/BACEN	Finanças públicas	BACEN
NFSP fluxo Governo	Finanças públicas	BACEN
NFSP fluxo BACEN	Finanças públicas	BACEN
NFSP fluxo estado	Finanças públicas	BACEN
NFSP fluxo município	Finanças públicas	BACEN
NFSP fluxo empresas estatais	Finanças públicas	BACEN
NFSP fluxo empresas federais	Finanças públicas	BACEN
NFSP fluxo empresas estaduais	Finanças públicas	BACEN
NFSP fluxo empresas municipais	Finanças públicas	BACEN
FOCUS1Media	Expectativa de mercado	BACEN
FOCUS1Mediana	Expectativa de mercado	BACEN
FOCUS1DP	Expectativa de mercado	BACEN
FOCUSTOP5media	Expectativa de mercado	BACEN
FOCUSTOP5mediana	Expectativa de mercado	BACEN

APÊNDICE B – RESULTADO DO TESTE DE ESTACIONARIDADE

Variável	Teste ADF	Teste ADF (primeira diferença)
IBC-Br	0.5887	0.01
Indicadores de produção industrial	0.0100	
Energia elétrica comercial	0.3070	0.01
Energia elétrica residencial	0.0100	
Energia elétrica outros	0.2414	0.01
Índice de vendas combustíveis	0.4618	0.01
Índice de vendas hipermercado	0.7070	0.01
Índice de vendas vestuário	0.0852	0.01
Índice de vendas eletrodomésticos	0.0500	
Índice de vendas automóveis	0.3313	0.01
Índice de vendas supermercado	0.8043	0.01
ICC	0.0480	0.01
ICEA	0.5074	0.01
IEF	0.5678	0.01
Exportação de bens	0.3216	0.01
Importação de bens	0.1027	0.01
Taxa CDB	0.7723	0.01
Taxa CDB pessoa física	0.0568	0.01
Taxa CDB pessoa jurídica	0.0637	0.01
Taxa CDB investidores institucionais	0.0477	
Taxa CDB pós-fixado	0.0879	0.01
TJLP	0.5548	0.01
CDI mensal	0.0351	
DLSP Governo	0.0399	
DLSP BACEN	0.9900	0.01
DLSP Estados	0.9830	0.01
DLSP Municípios	0.9603	0.01
DLSP Empresas federais	0.3620	0.01
DLSP Empresas estatais	0.0773	0.01
DLSP Empresas municipais	0.0539	0.01
IGPM	0.9900	0.01
IGPDI	0.0100	
IGP10	0.0100	
IPCA15	0.0100	
Agregado monetário M1	0.0100	
Agregado monetário M3	0.9694	0.01
Agregado monetário M4	0.9900	0.02221

Base monetária	0.9869	0.01
Empregos formais	0.9423	0.01
Empregos formais agricultura	0.9900	0.01
Empregos formais indústria	0.0100	
Empregos SIUP	0.9491	0.01
Empregos formais construção	0.9862	0.01
Empregos formais comércio	0.9749	0.01922
Vendas industriais	0.9900	0.01
Indicador indústria da transformação	0.1884	0.01
Horas trabalhadas da indústria	0.0100	
Utilização capacidade instalada	0.3221	0.01
Taxa de câmbio US\$ compra	0.1573	0.01
Taxa de câmbio US\$ venda	0.9445	0.01
Índice Ibovespa	0.9445	0.01
Operações compromissadas SELIC	0.6113	0.01
Títulos privados	0.4846	0.01
Quotas de fundos monetários	0.9900	0.01088
Títulos federais no poder público	0.9858	0.01
Desembolso BNDES	0.6029	0.01
Desembolso BNDES indústria	0.0100	
Desembolso BNDES comércio	0.0100	
Desembolso BNDES agricultura	0.0100	
Desembolso BNDES extração	0.0100	
Taxa SELIC mês	0.0114	
Dívida mobiliária Tesouro	0.0395	
Dívida mobiliária custódia Tesouro	0.9851	0.01
Dívida mobiliária custódia BACEN	0.9620	0.01
Dívida mobiliária saldo Tesouro	0.4324	0.01
Dívida mobiliária IGPDÍ carteira	0.9900	0.01
Dívida mobiliária IGPM carteira	0.4666	0.01
Dívida mobiliária TR	0.6547	0.01
Dívida mobiliária câmbio carteira	0.3344	0.01
Dívida mobiliária IGPDÍ custódia	0.0100	
Dívida mobiliária IGPM custódia	0.4666	0.01
Dívida mobiliária pré-fixada	0.7415	0.01
Dívida mobiliária câmbio custódia	0.9319	0.01
NFSP Governo/BACEN	0.0100	
NFSP Governo	0.7352	0.01
NFSP BACEN	0.7370	0.01
NFSP estado	0.0100	
NFSP município	0.0100	

NFSP empresas estatais	0.0100	
NFSP empresas federais	0.0100	
NFSP empresas estaduais	0.0100	
NFSP empresas municipais	0.0100	
NFSP fluxo Governo/BACEN	0.0100	
NFSP fluxo Governo	0.3009	0.01
NFSP fluxo BACEN	0.3043	0.01
NFSP fluxo estado	0.0100	
NFSP fluxo município	0.0100	
NFSP fluxo empresas estatais	0.0100	
NFSP fluxo empresas federais	0.0100	
NFSP fluxo empresas estaduais	0.0100	
NFSP fluxo empresas municipais	0.0100	
FOCUS1Media	0.0100	
FOCUS1Mediana	0.0100	
FOCUS1DP	0.0100	
FOCUSTOP5media	0.0100	
FOCUSTOP5mediana	0.0100	
FOCUSTOP5mediana	0.0100	
