

TEXTO PARA DISCUSSÃO

2814

PREVISÃO DE INFLAÇÃO: ANÁLISE  
PRELIMINAR DE DESEMPENHO DE  
TÉCNICAS DE *MACHINE LEARNING*

FRANCISCO LUNA SANTOS  
IZABEL NOLAU



**PREVISÃO DE INFLAÇÃO: ANÁLISE  
PRELIMINAR DE DESEMPENHO DE  
TÉCNICAS DE *MACHINE LEARNING***

**FRANCISCO LUNA SANTOS<sup>1</sup>  
IZABEL NOLAU<sup>2</sup>**

---

1. Técnico de planejamento e pesquisa na Diretoria de Estudos e Políticas Macroeconômicas do Instituto de Pesquisa Econômica Aplicada (Dimac/Ipea).

2. Pesquisadora do Subprograma de Pesquisa para o Desenvolvimento Nacional (PNPD) na Dimac/Ipea.

**Governo Federal**

**Ministério da Economia**

**Ministro** Paulo Guedes

**ipea** Instituto de Pesquisa  
Econômica Aplicada

Fundação pública vinculada ao Ministério da Economia, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

**Presidente**

**ERIK ALENCAR DE FIGUEIREDO**

**Diretor de Desenvolvimento Institucional**

**ANDRÉ SAMPAIO ZUVANOV**

**Diretor de Estudos e Políticas do Estado,  
das Instituições e da Democracia**

**FLÁVIO LYRIO CARNEIRO**

**Diretor de Estudos e Políticas Macroeconômicas**  
**MARCO ANTÔNIO FREITAS DE HOLLANDA CAVALCANTI**

**Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais**

**NILO LUIZ SACCARO JUNIOR**

**Diretor de Estudos e Políticas Setoriais de  
Inovação e Infraestrutura**

**JOÃO MARIA DE OLIVEIRA**

**Diretor de Estudos e Políticas Sociais**

**HERTON ELLERY ARAÚJO**

**Diretor de Estudos e Relações Econômicas e  
Políticas Internacionais**

**PAULO DE ANDRADE JACINTO**

**Coordenador-Geral de Imprensa e  
Comunicação Social (substituto)**

**JOÃO CLÁUDIO GARCIA RODRIGUES LIMA**

**OUVIDORIA:** <http://www.ipea.gov.br/ouvidoria>

**URL:** <http://www.ipea.gov.br>

## Texto para Discussão

Publicação seriada que divulga resultados de estudos e pesquisas em desenvolvimento pelo Ipea com o objetivo de fomentar o debate e oferecer subsídios à formulação e avaliação de políticas públicas.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2022

Texto para discussão / Instituto de Pesquisa Econômica  
Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1. Brasil. 2. Aspectos Econômicos. 3. Aspectos Sociais.  
I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos).  
Acesse: <http://www.ipea.gov.br/portal/publicacoes>

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

JEL: C01; C53.

DOI: <http://dx.doi.org/10.38116/td2814>

# SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO.....	6
2 METODOLOGIA .....	7
3 DADOS .....	16
4 RESULTADOS .....	18
5 TESTES ADICIONAIS.....	29
6 CONCLUSÃO .....	29
REFERÊNCIAS.....	31

## SINOPSE

Neste *Texto para Discussão* testamos diversos modelos de previsão de inflação e da atividade econômica brasileira, a partir de dados macroeconômicos e estimativas de especialistas, referentes ao período entre janeiro de 2002 e outubro de 2019, com frequência mensal. Devido à alta dimensão do conjunto de variáveis explicativas, nos concentramos em modelos associados à técnica de *machine learning* (ML) que oferecem diferentes abordagens para lidar com grandes conjuntos de dados e os comparamos com modelos de referência. Nossos resultados indicam que, para horizontes de previsão de curto prazo (um e três meses), os métodos de ML melhoram substancialmente as previsões de inflação. Para o horizonte de um mês, o Least Absolute Shrinkage and Selection Operator (Lasso) é o método com melhor desempenho, e a combinação dos melhores modelos de ML por horizonte é o destaque para o horizonte de três meses. No entanto, para horizontes de previsão mais longos (seis meses e doze meses), os modelos de ML e suas combinações não superam as expectativas de especialistas, ainda que as combinações de métodos de ML superem as metodologias tradicionais. Na previsão para o produto interno bruto (PIB), ocorre o contrário. Os modelos de ML não têm desempenho satisfatório para o horizonte mais curto, mas as combinações (horizonte de três e doze meses) e o *complete subset regression* (CSR) (horizonte de seis meses) superam os modelos tradicionais.

**Palavras-chave:** previsão; econometria; macroeconomia; *machine learning*.

## ABSTRACT

In this *Discussion Paper*, we test forecasting models for inflation and economic activity with macroeconomic data and economic surveys between January 2002 and October 2019 on a monthly basis. Due to the high dimension nature of the set of explanatory variables, we use machine learning (ML) models that offer different ways to deal with large datasets and we compare with benchmark models. We find that ML methods substantially improve inflation forecasts for shorter horizons (one and three months). While Least Absolute Shrinkage and Selection Operator (Lasso) is the model that best performs for the one-month horizon, a combination of ML models performs better for the three months horizon. However, for longer-term horizons (six and twelve months), individual ML methods and economic surveys do not perform well, despite the fact that a combination of ML models are better than benchmark models. Concerning GDP forecasts, the reverse is true. ML methods do not perform well for the one month horizon, but combinations of ML methods (three and twelve month) and complete subset regression (CSR) (six month) overcame traditional models.

**Keywords:** forecast; econometrics; macroeconomics; machine learning.

## 1 INTRODUÇÃO

Modelos de previsão de variáveis macroeconômicas são comumente utilizados como termômetro da atividade econômica e auxiliam no planejamento de ações de entes públicos bem como em decisões de agentes privados. No Brasil, o Banco Central do Brasil (BCB) disponibiliza o sistema de expectativas, que consolida previsões de mercado para um amplo conjunto de variáveis em frequência diária. Além de o sistema apresentar previsões para horizontes diversos, é possível selecionar a previsão somente das instituições com melhor desempenho nesse quesito. Muitos países da América Latina, como Chile, Colômbia e Peru, possuem sistemas de expectativas consolidados, em especial para variáveis relacionadas à política monetária.

Ainda que se tenham obtido avanços importantes, é primordial que se dê maior atenção ao tema da previsão, tendo em vista as dificuldades técnicas em obter boas estimativas. No caso do Brasil em particular, tais dificuldades são decorrência da volatilidade excessiva das taxas de inflação, num contexto de implantação recente, a partir de 1999, do sistema de metas de inflação. Essa volatilidade é transmitida e compartilhada por outras variáveis macroeconômicas, de forma que o objetivo de obter previsões se torna ainda mais desafiador do ponto de vista técnico. Em contrapartida, há diversas previsões feitas por especialistas disponíveis no mercado, o que facilita a comparação das diferenças técnicas e, no presente caso, a determinação se houve avanço com a introdução de técnica alternativa.

Neste estudo, o foco será a previsão de inflação e do produto interno bruto (PIB). A previsão de inflação é um dos pilares que norteiam a definição da taxa de juros básica em países com sistema de metas de inflação em que as expectativas assumem papel central. A variável de interesse, nesse caso, será o Índice Nacional de Preços ao Consumidor Amplo (IPCA) em frequência mensal.

Quanto ao PIB, ressalta-se sua importância para a definição dos parâmetros para o planejamento fiscal e orçamentário do setor público, bem como para a formação de expectativas no setor privado. Apesar de não haver uma métrica oficial mensal para o PIB, há indicadores bem-sucedidos que tentam replicá-la, como o monitor do PIB do Instituto Brasileiro de Economia da Fundação Getúlio Vargas (Ibre/FGV), que será nossa variável-alvo.

Com a disponibilidade cada vez maior de dados, crescem naturalmente as opções de em termos de modelagem. Nesse sentido, nos concentramos em modelos associados à técnica de *machine learning* (ML) que oferecem diferentes abordagens para lidar com grandes conjuntos de dados. Em ambientes ricos de dados, modelos de alta dimensão são necessários a fim de obter, entre as variáveis, padrões que permitam reduzir a dimensionalidade do problema utilizando o maior conjunto de informação possível.

Mais especificamente, os modelos de ML podem ser classificados em modelos de *shrinkage* (ou encolhimento) – Ridge, Least Absolute Shrinkage and Selection Operator (Lasso) e *adaptive* Lasso (adaLasso) –, modelos que combinam informações (*complete subset regression* – CSR), modelos de fatores (*principal component analysis* – PCA), modelos bayesianos (*Bayesian vector autoregressions* – BVAR) e árvores de regressão (*random forests*). Para fins de comparação, serão obtidas as previsões de nossos modelos de ML com uma série de modelos *benchmark* que incluem: modelos autorregressivos, *random walk* (RW), *exponentially weighted moving average* (EWMA). Além de constituírem variáveis em nossos modelos, as previsões dos especialistas compiladas pelo sistema *Focus* do BCB são igualmente utilizadas para comparar a eficácia dos modelos de ML especificamente para inflação.

Finalmente, em linha com Garcia, Medeiros e Vasconcelos (2017), uma característica que diferencia esta abordagem das demais é a utilização de dados em tempo real. Nesse sentido, o conjunto de informações utilizado na avaliação do desempenho das previsões engloba somente dados disponíveis no momento da previsão, eliminando potenciais interferências de dados previamente conhecidos que poderiam melhorar indevidamente os resultados.

O artigo está estruturado em mais cinco seções além desta introdução. Na seção 2, apresentamos as metodologias de ML utilizadas neste estudo. Na seção 3, descrevemos a base de dados e sua preparação para a execução das rotinas de ML. Na seção 4, apresentamos os resultados, seguidos de testes adicionais e das conclusões do trabalho na seção final.

## 2 METODOLOGIA

Por meio de ML, a ideia é construir algoritmos de modo a identificar padrões nos dados, não necessariamente relacionados a alguma teoria. Ao contrário da econometria, interpretação e inferência não são ingredientes essenciais. É particularmente útil a fim de desvendar aspectos ainda não compreendidos nos quais somente o aprendizado por máquina pode revelar novos padrões a partir dos dados. Além disso, permite mapear a economia ao longo de um número maior de dimensões ao tratar casos em que o número de variáveis pode, inclusive, ser maior que o número de observações. No Brasil, estudos iniciais que indicam que prever inflação por meio de ML tem desempenho superior aos modelos comumente utilizados.

Não há, portanto, seleção prévia de opções. O que se faz comumente na literatura de ML é testar os diversos modelos e, por meio de algum indicador de desempenho de previsão, escolher aquele que melhor prevê. Sendo assim, o desfecho de uma avaliação desse tipo pode resultar na escolha de modelos e variáveis explicativas diferentes para cada horizonte de tempo. Quanto ao problema de *overfitting*, os diferentes métodos de ML têm flexibilidade suficiente de forma

a permitir algum viés na estimação dos parâmetros, tendo como contrapartida o benefício da redução da variância da previsão.

Em seguida, apresentamos as principais características dos modelos, começando pelos modelos que servem como *benchmark* para, em seguida, explorar os modelos de ML.

## 2.1 Modelos *benchmark*

O primeiro modelo *benchmark* é o de passeio aleatório (RW), cuja representação é a soma do valor corrente da variável somada a um erro aleatório de média zero e variância constante. A característica principal do modelo RW é a persistência dos choques aleatórios, já que o efeito de cada choque não se dissipa, ou seja, tem memória infinita. Além disso, a melhor previsão para o período seguinte é justamente a variável-alvo corrente.

O segundo modelo corresponde ao modelo autorregressivo (AR) de ordem  $p$ , em que  $p$  é determinado pelo tradicional *Bayesian information criterion* (BIC) e os parâmetros são calculados por mínimos quadrados ordinários (*ordinary least squares* – OLS). Como o critério BIC pode indicar diferentes  $p$  para diferentes horizontes de tempo, os modelos podem ser diferentes para cada horizonte de tempo.

O terceiro *benchmark* é método de alisamento exponencial EWMA, que, por meio de um parâmetro de ponderação  $\lambda$ , permite que a variável-alvo mais recente tenha maior influência na previsão final. A ideia é que, ao impor pesos maiores para as observações mais recentes, será possível capturar melhor eventuais choques ou eventos extremos. Pode ser descrito como:

$$(1 - \lambda) \sum_{r=0}^{\infty} \lambda^r x_{t-r} \tag{1}$$

## 2.2 Métodos de ML

É fato que modelos lineares são surpreendentemente competitivos em relação aos métodos não lineares. No entanto, o modelo linear estimado via mínimos quadrados não é adequado para problemas de grande dimensionalidade por conta da variância dos erros de previsão. Nesta subseção, discutimos algumas maneiras pelas quais o modelo linear simples pode ser aprimorado, substituindo o ajuste de mínimos quadrados simples por procedimentos alternativos de ajuste.

Apesar de evidências apontarem para um bom poder preditivo dos modelos de ML, a escolha do melhor modelo de previsão é uma tarefa eminentemente prática, não há previsão teórica que

indique que um método seja superior ao outro em todas as situações. Nesse sentido, uma prática comum é combinar modelos como será demonstrando na seção 4.

Os métodos de ML utilizados neste texto são lineares e se enquadram no conceito de aprendizado supervisionado, em que a variável-alvo é conhecida. Em comum, esses métodos tornam possível lidar com mais variáveis do que observações e podem ser classificados em cinco tipos: métodos de *shrinkage*, que buscam reduzir a dimensionalidade do problema de previsão por meio do encolhimento dos parâmetros associados a variáveis irrelevantes; modelos que combinam informações de modelos encapsulados; métodos de seleção de fatores em que as variáveis são transformadas a fim de obter maior explicação possível da variância da resposta; modelos bayesianos; e árvores de regressão.

### 2.2.1 Métodos de *shrinkage*, ou encolhimento

Uma alternativa bem-sucedida para estimar parâmetros em grandes dimensões é usar métodos de *shrinkage* em que o sistema é ajustado com todos os preditores. Entretanto, os coeficientes estimados são encolhidos na direção do zero em relação às estimativas de mínimos quadrados. Esse encolhimento (conhecido como *regularização*) reduz a variância e, dependendo do método, alguns coeficientes podem ser exatamente iguais a zero.

A ideia dos diversos métodos é encolher para zero os parâmetros correspondentes a variáveis irrelevantes. Para tanto, ao processo usual de minimização de erros quadráticos é incorporado um termo adicional que é, em última instância, responsável por prover o melhor *trade-off* possível entre viés e variância. Ao restringir e/ou encolher os coeficientes estimados, podemos reduzir a variância com custo baixo em termos de viés.

Esse termo adicional inclui um parâmetro de penalização ( $\lambda$ ) cuja seleção é normalmente realizada por meio da estimação para diferentes valores de  $\lambda$  associada a algum critério de informação – Zou, Hastie e Tibshirani (2007) sugerem o BIC – ou de *cross-validation*.

A regressão Ridge, uma das principais opções em modelos de *shrinkage*, foi inicialmente proposta por Hoerl, Kennard e Baldwin (1975). Ridge minimiza a seguinte função:

$$\sum_{i=1}^T (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (2)$$

Em que  $\lambda$  é o parâmetro de penalização; e  $p$  é o número de variáveis.

Ridge tem ainda a vantagem de ter solução analítica. No entanto, ainda que os coeficientes estimados possam tender a zero, a geometria da penalização quadrática faz com que eles não sejam exatamente iguais a zero. O Ridge, portanto, não elimina variáveis irrelevantes tão somente minimizando seu impacto no modelo treinado. O parâmetro  $\lambda$  é escolhido entre uma sequência de valores possíveis por meio de *cross-validation*.

Mais recentemente, o Lasso, introduzido por Tibshirani (1996), e o *adaLasso*, de Zou (2006), receberam atenção especial entre os métodos de *shrinkage*. Em relação ao Ridge, a função de minimização do Lasso, mostrada a seguir, impõe penalização não linear nos coeficientes, o que faz com que o Lasso tenha a característica de selecionar as variáveis, não somente encolhê-las.

$$\sum_{i=1}^T (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

Os valores de  $\lambda$  que encolhem os coeficientes são limitados em ambos os lados,<sup>1</sup> o que facilita o problema do Lasso, porque sabemos onde procurar o melhor modelo e sem precisar estimar um número muito grande de modelos para encontrar uma boa solução em termos do BIC. Novamente, o parâmetro  $\lambda$  é escolhido entre uma sequência de valores possíveis por meio de *cross-validation*.

Apesar de trabalhar bem problemas de altíssima dimensão, como atestam Zhao e Yu (2006), o Lasso não possui a propriedade de oráculo (Zou, 2006), ou seja, não consegue selecionar o conjunto correto de variáveis relevantes sob determinadas condições. Para obter consistência na seleção de modelos, o Lasso requer uma condição bastante forte, denominada *irrepresentable condition*, que pressupõe baixa correlação entre variáveis relevantes e irrelevantes.

Nesse contexto, surgiu o método *adaLasso*, que se propõe a superar as deficiências do Lasso. Medeiros e Mendes (2017) mostraram que as condições que devem ser satisfeitas no *adaLasso* são mais gerais, e o modelo funciona mesmo quando o número de variáveis aumenta mais rapidamente que o número de observações e quando os erros são não gaussianos e heteroscedásticos. Trata-se de uma metodologia de duas etapas que usa um estimador do primeiro passo, geralmente o Lasso, para ponderar a importância relativa dos regressores. Sendo assim, o *adaLasso* é definido da seguinte forma:

$$\sum_{i=1}^T (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p w_j |\beta_j| \quad (4)$$

1. Entre o maior valor de  $\lambda$  que não causa encolhimento e o menor valor de  $\lambda$  que encolhe todos os coeficientes para zero.

## TEXTO para DISCUSSÃO

Em que  $w_j = (\beta_j^*)^{-\tau}$  representa diferentes pesos na penalização de cada variável, derivados do parâmetro  $\beta_j^*$  estimado em um primeiro passo. O parâmetro  $\tau$ , por sua vez, determina quanto queremos enfatizar a diferença dos pesos.

O mais usual é fazer  $\tau = 1$ . No entanto, Medeiros e Mendes (2017)<sup>2</sup> mostraram que a seleção desse parâmetro, usando o BIC da mesma forma que o  $\lambda$ , reduz os erros de previsão. No nosso caso, utilizamos  $\tau = 1$  e

$$w_j = \left| \beta_j^* + \frac{1}{\sqrt{T}} \right|^{-1} \quad (5)$$

Em que  $\beta_j^*$  representa a estimativa do coeficiente de cada variável obtida mediante um ajuste inicial do método Lasso.

Importante ressaltar novamente que o fato de o método somente encolher (Ridge) ou selecionar (Lasso e adaLasso) variáveis não determina inequivocamente qual método fornece a melhor previsão. O desempenho do modelo é uma questão empírica, e o encolhimento forçado a zero de variáveis, muito desejável do ponto de vista da análise e interpretação, pode eventualmente prejudicar a previsão, especialmente quando o conjunto de variáveis é mais homogêneo.

### 2.2.2 Modelo que combina informações: CSR

O modelo CSR, desenvolvido por Elliott, Gargano e Timmermann (2013; 2015), compõe uma classe diferente de método de ML. Ao contrário dos métodos de *shrinkage*, em que a ideia é minimizar uma função objetivo acrescida de um termo de penalização, o CSR consiste em avaliar todas as combinações possíveis tendo como base modelos lineares de pequena dimensão estimados a partir de um subconjunto de variáveis explicativas. Supondo que a base de dados seja composta por  $k$  variáveis independentes, o CSR pode ser estimado de acordo com as etapas que se seguem.

Defina o número  $k$  de variáveis em cada regressão linear, de forma que  $k < K$  (no nosso caso,  $k = 4$ ).

- 1) Estime todas as combinações possíveis de modelos com  $k$  variáveis.
- 2) Realize previsões para cada modelo.
- 3) A previsão final será a média das previsões do passo 3.

2. Eles se referem a esse modelo como Flex-adaLasso.

Apesar de simples, o CSR apresenta bons resultados de previsões e é robusto a *overfitting*. No entanto, uma importante restrição do modelo é que, para conjuntos grandes de variáveis explicativas, o número de regressões a serem estimadas pode ser muito grande, computacionalmente inviável.

De fato, este estudo lida com um número ainda maior de variáveis candidatas, de forma que será necessário adotar algum procedimento de pré-seleção de variáveis. Portanto, adotamos um procedimento de pré-teste, ajustando primeiro uma regressão linear em cada uma das variáveis candidatas – incluindo as defasagens – e salvando a estatística  $t$  de cada variável.

As estatísticas  $\tau$  são classificadas em valor absoluto e selecionamos as primeiras  $K = 20$  variáveis que são mais relevantes no *ranking*.<sup>3</sup> A previsão de CSR é calculada sobre essas variáveis. Todas as variáveis disponíveis são utilizadas no modelo.

### 2.2.3 Modelo de fatores

O objetivo de modelos de fatores é descrever a estrutura de covariância entre múltiplas variáveis por meio de um número reduzido de fatores latentes. Com isso, é possível captar características comuns de múltiplas séries temporais. Na prática, considere o modelo:

$$z_t = Lf_t + e_t \quad (6)$$

Em que  $(z_t, e_t)$  são matrizes de dimensão  $N \times T$ . Além disso,  $f_t$  tem dimensão  $q \times T$ , em que  $q \ll N$  e a matriz  $L$  tem dimensão  $N \times q$ .

Sob a hipótese de estacionariedade fraca e considerações adicionais sobre a estrutura de covariância<sup>4</sup> de  $(f_t, e_t)$ , a matriz de covariância de  $z_t$  ( $\Sigma_z$ ) pode ser escrita em função dos autovalores-autovetores correspondentes:

$$\Sigma_z = LL' + \Sigma_e = \lambda_1 a_1 a_1' + \lambda_2 a_2 a_2' + \dots + \lambda_N a_N a_N' \quad (7)$$

Considerando que os autovalores estão ordenados de forma decrescente, ou  $\lambda_1 > \lambda_2 > \dots > \lambda_N$ , se os primeiros  $q$  autovalores são grandes, podemos aproximar a matriz  $\Sigma_z$  como:

$$\Sigma_z \approx \lambda_1 a_1 a_1' + \lambda_2 a_2 a_2' + \dots + \lambda_q a_q a_q' \quad (8)$$

3. Uma alternativa seria usar o Lasso para selecionar as variáveis.

4. Basicamente:  $(f_t, e_t)$  tem média zero;  $\Sigma_e$  é uma matriz diagonal;  $\Sigma_f$  é matriz-identidade.

Assim, é possível reproduzir a estrutura de covariância das variáveis-alvo a partir de conjunto reduzido ( $q$ ) de variáveis latentes formadas a partir da combinação linear das variáveis originais. Assim como os modelos de *shrinkage*, modelos de fatores promovem a redução de dimensionalidade do problema econométrico e são apropriados para o caso em estudo.

### 2.2.4 Modelo bayesiano: BVAR

O modelo BVAR, popularizado por Doan, Litterman e Sims (1984) e Litterman (1986), pode ser considerado como um método de *shrinkage* multivariado. Num VAR tradicional, à medida que aumentamos o número de variáveis, o mesmo ocorre com a variação estimada dos parâmetros do modelo, levando a estimativas crescentemente imprecisas. Nesse sentido, os modelos BVAR também são usados para evitar o *overfitting*.

Considere o seguinte modelo:

$$y_t = c + \sum_{k=1}^p A_k y_{t-k} + \varepsilon_t \quad (9)$$

Em que  $y_t$  é um vetor  $n$ -dimensional composto pelas variáveis do VAR;  $c$  é um vetor de constantes;  $A_k$  é a matriz de coeficientes; e  $\varepsilon_t$  é vetor de erros gaussianos e matriz de covariância  $\Sigma_\varepsilon$ . Todas as variáveis são utilizadas no modelo.

A equação (9) pode ser reescrita como:

$$Y = XA + \varepsilon \quad (10)$$

Em que  $Y = (y_1, \dots, y_t)'$ ,  $X = (X_1, \dots, X_t)'$ ,  $y_t = (y_{1,t}, \dots, y_{n,t})'$ ,  $X_t = (1, y'_{t-1}, \dots, y'_{t-p})$ ,  $A = (c, A_1, \dots, A_p)$  e  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_t)$ .

A especificação da distribuição *a priori* leva em consideração que as defasagens mais recentes fornecem informações mais confiáveis do que as mais distantes; e que as próprias defasagens de uma determinada variável explicam mais da sua variação do que as defasagens de outras variáveis do modelo. As considerações anteriores são aplicadas no modelo a partir dos seguintes momentos da distribuição *a priori* dos coeficientes:

$$E(A_{k,i,j}) = \begin{cases} \delta_i, & j = i, k = 1 \\ 0, & \text{caso contrário} \end{cases} \quad \text{e} \quad V(A_{k,i,j}) = \begin{cases} \frac{\lambda^2}{k^2}, & j = i \\ \frac{\lambda^2 \sigma_i^2}{k^2 \sigma_j^2}, & \text{caso contrário} \end{cases} \quad (11)$$

Os coeficientes  $A_1, \dots, A_p$  são considerados independentes *a priori* e normalmente distribuídos. A matriz de covariância dos resíduos é considerada diagonal, fixa e conhecida:  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ . Finalmente, *a priori* do intercepto é difusa.

Originalmente, Litterman (1986) define  $\delta_i = 1$  para todo  $i$ , refletindo que todas as variáveis são caracterizadas por alta persistência, ou seja, supõe que os coeficientes são positivos. No entanto, essa *priori* não é apropriada para variáveis que se acredita serem caracterizadas por reversão substancial à média, isto é, oscilam em torno da média do modelo. Para esses, impomos uma *priori* ruído branco, definindo  $\delta_i = 0$ , como é o caso deste trabalho.

O parâmetro  $\lambda$  controla a importância relativa entre *a priori* e os dados: quando  $\lambda = 0$ , o modelo faz com que a distribuição *a posteriori* seja igual *a priori*, ignorando os dados; caso  $\lambda = 1$ , o modelo ignora *a priori* e temos estimativas iguais às de OLS. No nosso caso,  $\lambda = 0,1$ .

O fator  $1/k^2$  é a taxa na qual a variância *a priori* diminui com o aumento da defasagem e é responsável pelas diferentes escala e variabilidade dos dados.

Diferentemente dos métodos de *shrinkage*, a regularização<sup>5</sup> e a seleção de variáveis nos modelos BVAR são realizadas por meio da definição de distribuições *prior* independentes. Em um contexto de alta dimensão, o ideal é adotar um procedimento automático para definir as distribuições *prior*. Essa é exatamente a ideia subjacente ao Minnesota *prior* por Doan, Litterman e Sims (1984) e Litterman (1986), em que todas as médias das *priors* são definidas em zero, cuja aplicação em problemas de alta dimensionalidade é descrita em Banbura, Giannone e Reichlin (2010).

### 2.2.5 Árvores de regressão

Envolvem a estratificação ou *segmentação* do espaço de preditores em um número de regiões. Em seu conceito mais simples, para fazer uma projeção com árvores de regressão, usamos a média ou moda da observação de treinamento da região a que a nova observação pertence. Métodos de árvore são simples e úteis para interpretação, mas não têm desempenho comparável ao dos melhores métodos supervisionados em termos de acurácia de previsão. Sendo assim, quanto maior a complexidade da árvore de decisão, maior a acurácia e menor a interpretabilidade.

Seja  $x_i = (x_{i,1} \dots x_{i,m})'$  um conjunto de  $m$  variáveis explicativas da variável-alvo  $y_i$ , suponha que a relação entre  $x_i$  e  $y_i$  possa ser descrita por uma função  $f(x_i)$  tal que:

5. Ou seja, evitar o *overfitting*.

## TEXTO para DISCUSSÃO

$$y_i \approx f(x_i) + \varepsilon_i \quad (12)$$

Uma regressão em árvore com  $K$  nós terminais tenta aproximar a função  $f$  por uma função  $H(x_i, \psi)$ , não linear, em que  $\psi$  é um vetor de parâmetros.  $H(\cdot)$  é uma função por partes com  $K$  sub-regiões que são ortogonais ao eixo das variáveis explicativas.

Cada sub-região representa um nó terminal  $k_j(\theta_j)$ , para  $j = 0, \dots, K-1$ . O parâmetro  $\theta_j$  define cada sub-região tal que:

$$f(x_i) = \sum_{j=0}^{K-1} \beta_j \cdot I_j(x_i, \theta_j) \quad (13)$$

Em que  $I_j(x_i, \theta_j)$  é uma função indicador tal que:

$$I_j(x_i, \theta_j) = \begin{cases} 1, & \text{se } x_i \in k_j(\theta_j) \\ 0, & \text{caso contrário} \end{cases} \quad (14)$$

Cada nova partição na árvore é criada por meio de um problema de otimização cujo objetivo é achar a partição que tem a maior contribuição na redução do erro quadrático do modelo. Para isso, o processo de busca pela partição ótima consiste em avaliar todas as possíveis variáveis e observações. Árvores de regressão, no entanto, são conhecidas por apresentar grande instabilidade, ou seja, qualquer mudança nos dados pode acarretar uma grande mudança do modelo.

Dentro dessa abordagem, o método de *random forest*, cujo objetivo é retirar a correlação entre as diversas áreas da árvore, se destaca pela redução na variância dos erros de previsão. Trata-se de um algoritmo que combina regressão em árvore com técnicas de agregação por *bootstrap (bagging)*.

A estimação segue as etapas a seguir:

- gerar uma amostra  $b$  via *bootstrap* dos dados com reposição, sendo que normalmente  $b$  tem o mesmo número de observações da amostra;
- na amostra aleatória, construir uma árvore de regressão; e
- repetir os passos anteriores  $B = 500$  vezes e calcular a previsão de acordo com a média das previsões em todas as árvores estimadas.

Em cada decisão de abertura da árvore (decisão de separação), somente  $m$  variáveis (das  $k$  variáveis possíveis) são consideradas. Além disso, as  $m = p/3$  variáveis candidatas são aleatórias a cada passo da árvore.

### 3 DADOS

O modelo de previsão usou dados macroeconômicos e previsões de especialistas do período entre janeiro de 2002 e outubro de 2019, com frequência mensal. Modelos de alta dimensionalidade, os de ML em particular, funcionam bem quando o conjunto de dados tem qualidade. Caso contrário, há risco de escolha de variáveis irrelevantes em detrimento das relevantes.

Os dados macroeconômicos do trabalho foram reunidos a partir de várias fontes públicas, como BCB, Instituto Brasileiro de Geografia e Estatística (IBGE) e Ipeadata, além dos dados internacionais do Federal Reserve Economic Data (Fred). Além disso, há variáveis cuja fonte é privada, notadamente os dados de confiança, índice de utilização de capacidade e o monitor do PIB da FGV; e índices de risco e preços de *commodities* da Bloomberg. Ao todo, há 219 variáveis macroeconômicas que, após a coleta, passam por etapa não trivial de preparação para que os dados sejam empregados na previsão, como será explicado na subseção seguinte.

Finalmente, as previsões dos especialistas, que representam as expectativas de mercado para o IPCA, foram obtidas no *Focus*. O relatório, criado para apoiar o regime de metas de inflação, é publicado pelo BCB a cada semana, embora as coletas com agentes no mercado sejam feitas diariamente. Utilizamos a mediana dessas expectativas de mercado para o mês seguinte e para os doze seguintes.

#### 3.1 Preparação dos dados

O conjunto de dados é composto pelas 219 variáveis macroeconômicas coletadas, além de uma variável indicadora de mês para os meses de janeiro a novembro, totalizando onze variáveis *dummies*. Essas variáveis assumem valor um, se o mês de referência corresponde ao mês da variável *dummy*, e valor zero, caso contrário. Note que não há variável *dummy* associada ao mês de dezembro, já que isso geraria um problema de multicolinearidade. As defasagens da variável-alvo também são introduzidas ao conjunto de variáveis explicativas, ou seja, acrescentamos o deslocamento da variável IPCA ou PIB no tempo. Optamos por utilizar três defasagens da variável-alvo. No mais, incorporamos à base de dados as expectativas (*exp*) de mercado para o IPCA dadas pela média, obtidas no relatório *Focus*, para o mês de referência

## TEXTO para DISCUSSÃO

e os doze meses seguintes, totalizando treze variáveis. Por fim, temos então uma base de dados com 246 variáveis e 214 tempos (janeiro de 2002 a outubro de 2019).

Um dos problemas dessa base de dados são os valores faltantes, que podem ocorrer no início da série (sem valores antigos disponíveis) ou no fim (sem valores recentes disponíveis). Para lidar com essa questão, cada uma das séries que têm essas características, exceto a variável-alvo, é deslocada temporalmente até que não haja mais valores faltantes no seu fim. Note que esses valores faltantes são remanejados para o início da série. No caso em que há valores faltantes no fim da variável-alvo, os tempos referentes a esses valores faltantes são extintos e, conseqüentemente, perdemos informação de quaisquer variáveis naquele instante e nos seguintes.

Agora, temos uma base de dados completamente preenchida nos últimos tempos, mas com valores faltantes no início. Inclusive, algumas variáveis possuem mais valores faltantes do que observações. Para resolver esse problema, vamos encontrar um *trade-off* entre a quantidade de variáveis e a quantidade de tempos a serem removidos. A tabela 1 mostra a relação entre essas duas quantidades, considerando o IPCA como variável-alvo.

**TABELA 1**

**Relação entre quantidade de variáveis e quantidade de tempos a serem removidos**

Tempos a serem removidos ( $t$ )	0	1	2	3	11	13	108	109
Variáveis a serem removidas ( $v$ )	149	41	23	22	11	6	4	0

Elaboração dos autores.

Note que, quanto maior o número de tempos a serem removidos, menos variáveis devemos tirar da nossa base, já que as observações faltantes estão no início das séries. O critério utilizado para escolher a quantidade de tempos a serem removidos é avaliar a diferença entre os tempos: quando a diferença  $t_{i+1} - t_i$  entre os tempos a serem removidos for maior que sete, eliminamos  $t_i$  tempos da nossa base e  $v_i$  variáveis. Note que, se removêssemos de três a dez tempos, estaríamos eliminando 22 variáveis da base; e se removêssemos onze tempos, estaríamos eliminando onze variáveis da base. Assim, quando a variável-alvo é o IPCA, eliminamos 3 tempos e 22 variáveis da base, resultando num conjunto de dados final com 211 tempos (abril de 2002 a outubro de 2019) e 224 variáveis, sem dados faltantes. No caso do PIB, temos um conjunto de dados final com 209 tempos (maio de 2002 a setembro de 2019) e as mesmas 224 variáveis.

## 4 RESULTADOS

Nesta seção, descrevemos os métodos usados para prever a variável-alvo (PIB ou inflação)  $h$  períodos à frente ( $y_{t+h}$ ) no tempo  $t$ . A medida de inflação que se quer prever é o IPCA, que é um dos índices mais importantes da nossa economia e serve como parâmetro para as decisões de política monetária. Quanto ao PIB, utilizamos a medida mensal do monitor do PIB da FGV, que tem como base a mesma metodologia das Contas Nacionais do IBGE.

Como em Medeiros *et al.* (2018), consideramos uma abordagem de previsão direta, em que  $y_{t+h}$  é modelado como uma função de um conjunto de variáveis explicativas  $x_t$  medido no tempo  $t$ , como:

$$y_{t+h|t} = T(x_t) + \mu_{t+h} \quad (15)$$

Em que  $T(x_t)$  corresponde ao mapeamento entre o conjunto de variáveis explicativas e a variável-alvo futura; e  $\mu_{t+h}$  é o erro de previsão.

Na abordagem direta não é feita previsão das variáveis explicativas, ou seja, utiliza-se o último valor observado disponível. A exceção é o modelo BVAR que calcula previsões para  $x_t$ , conforme descrito em Banbura, Giannone e Reichlin (2010). Nesse caso, todas as variáveis são previstas, mas utilizamos apenas a previsão da variável de interesse  $y$ . As previsões acumuladas são calculadas por meio da agregação das previsões individuais para cada  $t$ . É importante ressaltar que as previsões não são em tempo real já que variáveis divulgadas com defasagem fazem parte do conjunto de informações  $x_t$ .

Em geral, não nos importamos com a eficácia do método na amostra de treinamento já que não há garantia de que um bom ajuste dentro da amostra seja extensivo ao período fora da amostra. O sobreajuste (*overfitting*) ocorre quando o modelo ( $T(x_t)$ ) tenta replicar supostos padrões presentes nos dados de treinamento que simplesmente não existem nos dados de teste. No contexto de previsão,  $T(x_t)$  é tratado como caixa-preta, ou seja, a preocupação não é recuperar sua forma exata desde que gere previsões acuradas de  $y_{t+h|t}$ . A acurácia de  $T(x_t)$  como preditor para  $y_{t+h|t}$  depende de duas quantidades: um erro redutível e um erro irredutível.

Suponha que exista uma função desconhecida  $f(x_t)$ , tal que  $y_{t+h|t} = f(x_t) + \varepsilon_t$ . Nosso objetivo é encontrar a melhor função  $T(x_t)$  ainda que ela não seja uma estimativa perfeita para a função  $f(x_t)$ . O erro associado a essa busca é chamado de redutível, já que podemos melhorar a aderência de  $T(x_t)$  usando a técnica de aprendizado estatístico mais apropriada para recuperar a melhor especificação. No entanto, mesmo que fosse possível formar uma estimativa perfeita para  $f(x_t)$ , a previsão ainda apresentaria algum erro. Isso ocorre porque  $y_{t+h|t}$  também é uma função do erro  $\varepsilon$ , que, por definição, não pode ser previsto usando  $x_t$ . Portanto, a variabilidade

## TEXTO para DISCUSSÃO

associada a  $\varepsilon$  também afeta a precisão de nossas previsões. Este é o erro irreduzível, porque, por mais que nos aproximemos  $f(x_t)$ , não podemos reduzir o erro introduzido por  $\varepsilon$ .

Para um modelo específico  $T(\cdot)$ , o *mean squared error* (MSE) em particular pode ser decomposto em termos dos erros redutível e irreduzível da seguinte forma:

$$E(y_0 - T(x_0))^2 = \underbrace{Var(T(x_0)) + [Viés(T(x_0))]^2}_{\text{redutível}} + \underbrace{Var(\varepsilon)}_{\text{irreduzível}} \quad (16)$$

Em que  $y_0$  é a observação fora da amostra;  $T(x_0)$  é a previsão fora da amostra; e  $Var(\varepsilon)$  é a variância do erro irreduzível. A notação  $E(y_0 - T(x_0))$  define o valor esperado do teste MSE e se refere à média do teste MSE que seria obtida se repetíssemos a estimativa de  $f$  usando um número grande de amostras de teste e testássemos a cada  $x_0$  da amostra de teste.

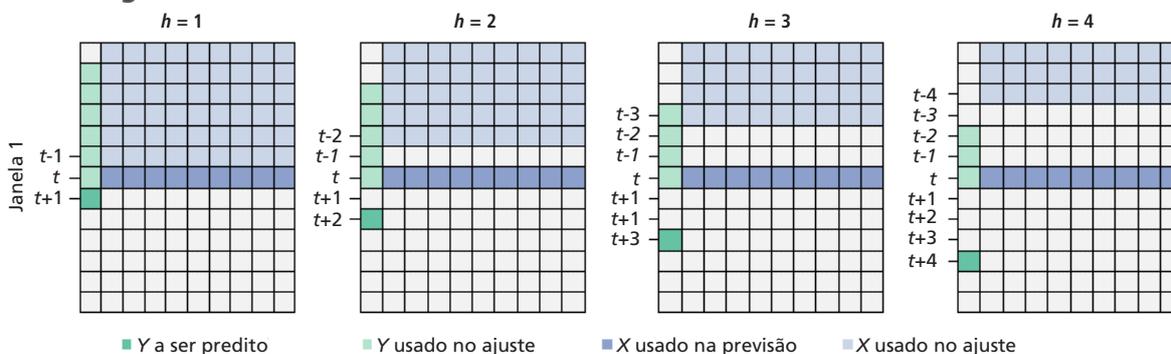
A busca pela função ótima, portanto, pode ser resumida em termos do *trade-off* entre viés e variância do modelo. O método de aprendizagem precisa achar um modelo que atinja baixa variância e baixo viés. Variância se refere à quantidade que a previsão  $T(x_0)$  mudaria se estimássemos novamente com uma diferente amostra de treinamento. Viés, por sua vez, se refere ao erro de "simplificação" de um problema complexo em um modelo mais simples e tratável. A taxa de mudança relativa dessas duas quantidades determina se o MSE de teste aumenta ou diminui. Aumento de flexibilidade leva a uma redução relativa do viés maior do que o aumento da variância. Nesse caso, o MSE de teste esperado cai. No entanto, a redução marginal do viés é decrescente, de forma que há um ponto de convergência nesse *trade-off*.

Dada uma amostra, resta definir a melhor maneira de realizar o teste fora da amostra, ou *cross-validation*. A literatura apresenta opções de amostragem aleatória (*k-fold*, *leave-one-out*, entre outros) que não atendem ao objetivo porque se apoiam de forma crucial da hipótese de erros independentes e identicamente distribuídos, o que ignoraria as autocorrelações presentes na estrutura temporal das séries.

No trabalho, portanto, foi utilizada uma configuração de testes com janelas deslizantes (*rolling window*). A figura 1 apresenta uma ilustração de como o *rolling window* funciona, tendo, na primeira coluna, a variável-alvo, denotada por  $Y$ , e nas outras colunas as variáveis explicativas, denotadas por  $X$ .

FIGURA 1

## Rolling window



Elaboração dos autores.

No exemplo, ilustramos como o *rolling window* faz previsões para  $h = 1, \dots$ , passos à frente, considerando a sua primeira janela. Na previsão 1 passo à frente, ajustamos os modelos considerando as variáveis explicativas no intervalo  $[1; t - 1]$ , denotadas por  $X_{1:t-1}$ , e considerando a variável-alvo no intervalo  $[2; t]$ . Assim, nesse caso, os modelos ajustados resultantes nos fornecerão previsões 1 passo à frente, e, conseqüentemente, a previsão da variável-alvo será dada por  $\hat{Y}_{t+1} = T(X_t)$ . Para fazer previsões  $h$  passos à frente, ajustamos os modelos considerando as variáveis explicativas no intervalo  $[1; t - 1]$  e a variável-alvo no intervalo  $[1 + h; t]$ . Assim, os modelos ajustados resultantes nos fornecerão previsões  $h$  passos à frente, e a previsão da variável-alvo será dada por  $\hat{Y}_{t+h} = T(X_t)$ . Para a previsão acumulada do mês  $h$ , consideramos a seguinte equação:

$$100 \left( \prod_{i=1}^h \left( 1 + \frac{\hat{Y}_i}{100} \right) - 1 \right) \quad (17)$$

Neste trabalho, utilizamos uma configuração do *rolling window* considerando 48 janelas de tamanho 163 para o IPCA e 161 para o PIB. As previsões foram calculadas com base nessas janelas rolantes e computadas para  $h = 1, \dots$ , meses à frente. Neste relatório apresentamos a previsão de um mês à frente e as previsões acumuladas de três, seis e doze meses à frente. Para cada uma das 48 janelas do *rolling window* temos doze previsões (um a doze meses à frente). Podemos então validar os modelos comparando as previsões com os seus respectivos valores verdadeiros.<sup>6</sup>

Além dos modelos propostos, foram sugeridas algumas combinações de suas previsões. As combinações apresentadas a seguir combinam previsões dos modelos para cada um dos doze

6. Note que as previsões das últimas janelas (condicional a  $h$ ) não podem ser comparadas, já que as observações estão fora da amostra.

horizontes e, posteriormente, essas previsões são acumuladas de três, seis e doze meses à frente, para fins. São elas as que se seguem.

- 1) *Model confidence set* (MCS) é um teste que foi desenvolvido por Hansen, Lunde e Nason (2011), que retorna um conjunto que contém o melhor modelo com probabilidade  $100(1 - \alpha)\%$ . Quanto maior o valor de  $\alpha$ , há uma tendência de redução no número de modelos do conjunto e vice-versa. O MCS permite comparar vários modelos de forma conjunta, ao contrário do teste de Diebold-Mariano, que só permite fazê-lo em duplas. Neste trabalho, usamos o MCS não para encontrar um modelo específico, mas como ferramenta de combinação de modelos. A escolha dos modelos ficou restrita aos de ML, e a previsão final é a média dos modelos selecionados no MCS.
- 2) Modelo por horizonte (MPH): a partir do MSE, escolhemos o melhor MPH e consideramos suas previsões.
- 3) MPH-ML: a partir do MSE, escolhemos o melhor modelo entre os de ML por horizonte e consideramos suas previsões.

Já as combinações a seguir avaliam as previsões um mês à frente e acumuladas de três, seis e doze meses à frente.

- 1) Média: toma a média das previsões dos modelos.
- 2) Média-ML: toma a média das previsões dos modelos entre os de ML.
- 3) Mediana: toma a mediana das previsões dos modelos.
- 4) Média-trim: toma a média desconsiderando 25% das observações de cada extremidade das previsões dos modelos.

Para medir a qualidade do ajuste de previsão e, assim, determinar o melhor modelo, optamos por utilizar quatro diferentes métricas. Na métrica *mean absolute error* (MAE), a função perda é o erro de previsão absoluto, enquanto MSE e *root mean squared error* (RMSE) utilizam a função perda quadrática como referência. Ao contrário das demais, a estatística U de Theil (Theil-U) é uma medida relativa de erro de previsão em que o desempenho preditivo de cada modelo é comparado ao de um modelo simples, no caso o passeio aleatório.

Os resultados são reportados nas tabelas 2 e 3 para todos os modelos propostos e suas combinações, considerando as previsões um mês à frente e acumuladas de três, seis e doze meses à frente. Adicionalmente, para cada horizonte de previsão, reportamos o modelo econométrico mais acurado e, no caso do IPCA, apresentamos também as expectativas do mercado para o IPCA dadas pela média, obtidas no relatório *Focus*.

TABELA 2

IPCA: MAE, MSE, RMSE e Theil-U referentes aos modelos propostos e suas combinações

Horizonte		T+1				T+3			
Métrica		MAE	MSE	RMSE	Theil-U	MAE	MSE	RMSE	Theil-U
	EWMA	0,2316	0,0954	0,3088	0,9829	0,6515	0,6973	0,8351	0,9567
	BVAR	0,2018	0,0787	0,2805	0,9368	0,6054	0,5714	0,7559	0,9103
	RW	0,2446	0,1022	0,3196	1,0000	0,6780	0,8323	0,9123	1,0000
	Lasso	0,1958	0,0746	0,2731	0,9244	0,5122	0,3876	0,6226	0,8261
	Ridge	0,2041	0,0763	0,2763	0,9297	0,5585	0,4319	0,6572	0,8487
	CSR	0,1984	0,0750	0,2739	0,9258	0,5547	0,4313	0,6567	0,8484
	AdaLasso	0,1959	0,0775	0,2783	0,9332	<b>0,4899</b>	0,3750	0,6124	0,8193
	Random forest	0,2186	0,0873	0,2955	0,9565	0,6213	0,5195	0,7208	0,8889
	PCA	0,2474	0,1096	0,3311	0,9897	0,6146	0,5665	0,7527	0,9082
	AR	0,2084	0,0812	0,2849	0,9441	0,5939	0,5098	0,7140	0,8847
	MCS-ML	0,2018	0,0787	0,2805	0,9368	0,5509	0,4723	0,6873	0,8683
	MPH	0,1958	0,0746	0,2731	0,9244	0,4940	<b>0,3676</b>	<b>0,6063</b>	<b>0,8152</b>
	MPH-ML	0,1958	0,0746	0,2731	0,9244	0,4940	<b>0,3676</b>	<b>0,6063</b>	<b>0,8152</b>
Combinações	Média	<b>0,1951</b>	0,0752	0,2743	0,9263	0,5281	0,4132	0,6428	0,8404
	Média-ML	0,1953	<b>0,0739</b>	<b>0,2719</b>	<b>0,9220</b>	0,5432	0,4155	0,6446	0,8405
	Mediana	0,1955	0,0754	0,2746	0,9268	0,5248	0,4034	0,6351	0,8363
	Média-trim	0,1961	0,0745	0,2729	0,9237	0,5237	0,3978	0,6307	0,8326
	Focus-média	0,2285	0,0887	0,2978	0,9652	0,4940	0,3836	0,6193	0,8239
Horizonte		T+6				T+12			
Métrica		MAE	MSE	RMSE	Theil-U	MAE	MSE	RMSE	Theil-U
	EWMA	1,2634	2,9496	1,7175	0,9495	2,6725	14,2188	3,7708	0,9628
	BVAR	1,2740	2,2759	1,5086	0,8899	2,8690	12,8402	3,5833	0,9385
	RW	1,3636	3,6292	1,9050	1,0000	2,8446	16,5499	4,0682	1,0000
	Lasso	0,9991	1,4244	1,1935	0,7915	2,2298	7,4764	2,7343	0,8198
	Ridge	1,0654	1,6795	1,2959	0,8248	2,6532	11,0999	3,3316	0,9050
	CSR	1,0633	1,6658	1,2906	0,8231	2,6095	10,5307	3,2451	0,8931
	AdaLasso	0,9374	1,3685	1,1698	0,7836	2,3231	7,4737	2,7338	0,8198
	Random forest	1,1383	1,7261	1,3138	0,8307	2,5132	9,1655	3,0275	0,8623
	PCA	1,2375	2,3982	1,5486	0,9035	3,1903	15,4753	3,9339	0,9822
	AR	1,0123	1,4909	1,2210	0,8006	1,8346	<b>4,7605</b>	<b>2,1819</b>	<b>0,7323</b>
	MCS-ML	1,0658	1,6250	1,2747	0,8181	2,6213	10,8723	3,2973	0,9003
	MPH	<b>0,9135</b>	<b>1,2169</b>	<b>1,1031</b>	<b>0,7610</b>	<b>1,7994</b>	4,7737	2,1849	0,7328
	MPH-ML	0,9878	1,3228	1,1501	0,7766	2,1912	6,8858	2,6241	0,8021
Combinações	Média	1,0080	1,4672	1,2113	0,7968	2,3978	8,7169	2,9524	0,8517
	Média-ML	1,0430	1,5612	1,2495	0,8099	2,4830	9,2012	3,0333	0,8634
	Mediana	0,9822	1,3604	1,1664	0,7805	2,3569	8,3808	2,8950	0,8421
	Média-trim	0,9907	1,4013	1,1838	0,7867	2,3557	8,2496	2,8722	0,8401
	Focus-média	<b>0,7338</b>	<b>0,9112</b>	<b>0,9546</b>	<b>0,7079</b>	<b>1,4992</b>	<b>3,4978</b>	<b>1,8702</b>	<b>0,6780</b>

Elaboração dos autores.

## TEXTO para DISCUSSÃO

### TABELA 3

PIB: MAE, MSE, RMSE e Theil-U referentes aos modelos propostos e suas combinações

Horizonte		T+1				T+3			
Métrica		MAE	MSE	RMSE	Theil-U	MAE	MSE	RMSE	Theil-U
	EWMA	0,5220	<b>0,8313</b>	<b>0,9117</b>	<b>0,7833</b>	0,7128	1,2720	1,1278	0,5930
	BVAR	0,5710	0,9402	0,9696	0,8078	0,9052	1,6440	1,2822	0,6323
	RW	0,8486	2,2082	1,4860	1,0000	1,8148	10,2847	3,2070	1,0000
	Lasso	0,5334	0,8362	0,9144	0,7844	0,8601	1,3263	1,1517	0,5993
	Ridge	0,5349	0,8592	0,9269	0,7898	0,6881	1,3423	1,1586	0,6011
	CSR	<b>0,5077</b>	0,8429	0,9181	0,7860	0,6343	1,1996	1,0953	0,5844
	AdaLasso	0,5567	0,8479	0,9208	0,7872	0,9684	1,5942	1,2626	0,6275
	<i>Random forest</i>	0,6097	1,2439	1,1153	0,8684	0,8406	2,5124	1,5851	0,6987
	PCA	0,6181	0,9614	0,9805	0,8229	0,9026	1,7908	1,3382	0,6415
	AR	0,5720	0,9549	0,9772	0,8109	0,8496	1,4496	1,2040	0,6127
	MCS-ML	0,5710	0,9402	0,9696	0,8078	0,7653	1,5124	1,2298	0,6180
	MPH	0,5220	<b>0,8313</b>	<b>0,9117</b>	<b>0,7833</b>	<b>0,6178</b>	<b>1,1772</b>	<b>1,0850</b>	<b>0,5817</b>
	MPH-ML	0,5334	0,8362	0,9144	0,7844	0,6986	1,1857	1,0889	0,5827
Combinações	Média	0,5424	0,9540	0,9767	0,8106	0,7626	1,6561	1,2869	0,6323
	Média-ML	0,5340	0,8924	0,9447	0,7978	0,7213	1,3799	1,1747	0,6047
	Mediana	0,5278	0,8627	0,9288	0,7880	0,6620	1,2263	1,1074	0,5885
	Média-trim	0,5223	0,8685	0,9319	0,7900	0,6787	1,2775	1,1303	0,5931
Horizonte		T+6				T+12			
Métrica		MAE	MSE	RMSE	Theil-U	MAE	MSE	RMSE	Theil-U
	EWMA	1,1218	2,3912	1,5463	0,5083	2,1923	7,3986	2,7200	0,4693
	BVAR	1,2467	2,7773	1,6665	0,5277	2,1023	8,0559	2,8383	0,4793
	RW	3,4546	35,8100	5,9841	1,0000	6,9632	152,593	12,3528	1,0000
	Lasso	1,1840	2,2584	1,5028	0,5011	1,6772	4,4157	2,1013	0,4124
	Ridge	1,0065	2,0424	1,4291	0,4887	1,7608	5,1917	2,2785	0,4295
	CSR	<b>0,8560</b>	<b>1,6953</b>	<b>1,3020</b>	<b>0,4665</b>	1,7709	5,6140	2,3694	0,4380
	AdaLasso	1,3966	2,9689	1,7231	0,5366	1,9778	5,4621	2,3371	0,4350
	<i>Random forest</i>	1,2405	5,0323	2,2433	0,6109	2,0642	14,5248	3,8111	0,5565
	PCA	1,2757	2,8643	1,6924	0,5223	1,7455	4,9723	2,2299	0,4065
	AR	1,3279	2,7068	1,6452	0,5243	1,6870	5,2345	2,2879	0,4304
	MCS-ML	1,0053	2,3279	1,5258	0,5044	1,6390	5,6079	2,3681	0,4381
	MPH	0,8975	1,8026	1,3426	0,4768	1,2843	3,1700	1,7805	0,3738
	MPH-ML	0,9427	1,7828	1,3352	0,4724	<b>1,2195</b>	<b>2,7806</b>	<b>1,6675</b>	<b>0,3674</b>
Combinações	Média	1,0979	2,8162	1,6781	0,5291	1,8358	7,0129	2,6482	0,4621
	Média-ML	0,9459	2,0185	1,4207	0,4871	1,4110	4,1131	2,0281	0,4057
	Mediana	0,9216	1,8283	1,3521	0,4755	1,5760	4,3939	2,0962	0,4080
	Média-trim	0,9237	1,8531	1,3613	0,4781	1,4801	3,9652	1,9913	0,3983

Elaboração dos autores.

Os modelos associados a menores valores de MAE, MSE, RMSE e Theil-U, dados em negrito, fornecem melhores previsões fora da amostra. A partir da tabela 2, podemos perceber que, para o horizonte  $T+1$ , as melhores previsões entre os modelos propostos foram obtidas por meio das combinações média (MAE) e média-ML (para os critérios restantes). Nos horizontes  $T+3$  e  $T+6$ , a combinação MPH nos fornece menores valores dos critérios utilizados, exceto o critério MAE do horizonte  $T+3$ , em que o melhor modelo é o *adaLasso*. No mais, no horizonte  $T+12$ , o modelo AR é indicado como melhor de acordo com o MSE, RMSE e Theil-U, enquanto o MAE indica a combinação MPH como melhor. Nos horizontes  $T+6$  e  $T+12$ , os modelos escolhidos perdem para o *Focus*. Note que, no horizonte  $T+3$ , os valores dos critérios são os mesmos para MPH e MPH-ML, indicando que pelo menos nos três primeiros horizontes os melhores modelos de acordo com o MSE foram dados por modelos de ML.

Os resultados apresentados na tabela 3 indicam que, para os horizontes  $T+1$  e  $T+3$ , as melhores previsões foram obtidas pela combinação MPH, embora, no primeiro horizonte, o MPH seja dado pelo modelo EWMA. No entanto, CSR e MPH-ML, que envolvem o modelo de ML, são os melhores métodos para os horizontes mais longos, seis e doze meses, respectivamente. Note que, horizontes mais longos são caracterizados por variações maiores e, conseqüentemente, erros maiores.

Para cada horizonte de previsão, os gráficos 1 e 2 apresentam os resultados do nosso exercício de previsões um mês à frente e acumuladas de três, seis e doze meses à frente para o IPCA e o PIB, respectivamente, considerando apenas os métodos de ML. O eixo que contém as datas corresponde ao momento que estamos prevendo. Por exemplo, se algum ponto mostrar janeiro de 2015, isso significa que a previsão de um passo à frente é a previsão para esse mês; e a previsão cumulativa de três meses é para a inflação acumulada de janeiro a março.

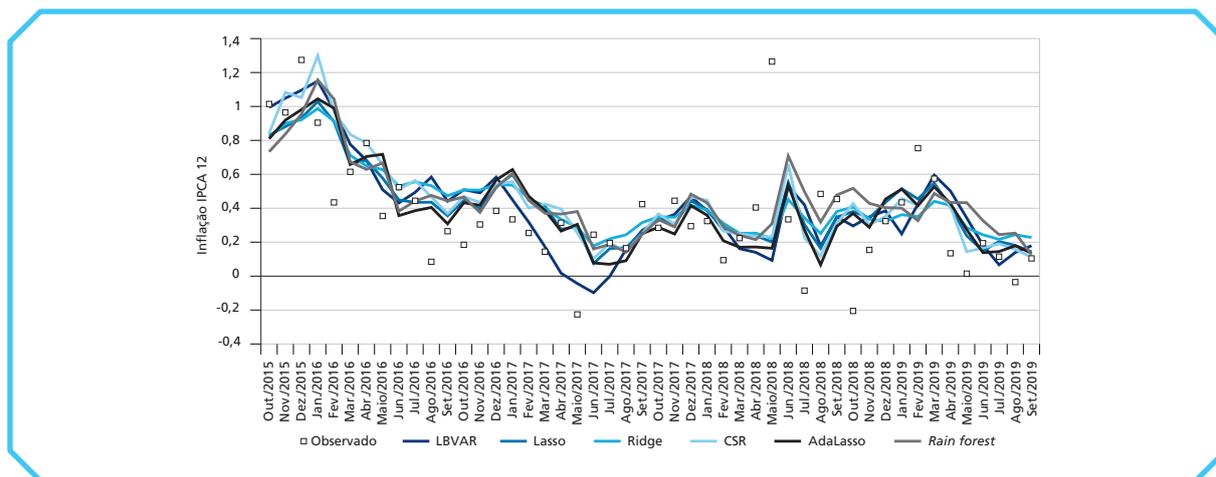
A partir do gráfico 1, notamos que, para horizontes curtos, todos os modelos estão próximos e apresentam comportamentos semelhantes, como uma versão mais suave da inflação observada, pois perdem alguns altos e baixos nos registros, postergando suas reversões. Já para horizontes mais longos, apesar dos comportamentos semelhantes, podemos perceber que o *adaLasso* tende a ficar por baixo dos demais, enquanto o CSR (até o fim de 2017) e o *random forest* (a partir de 2018) tendem a ficar por cima.

TEXTO para DISCUSSÃO

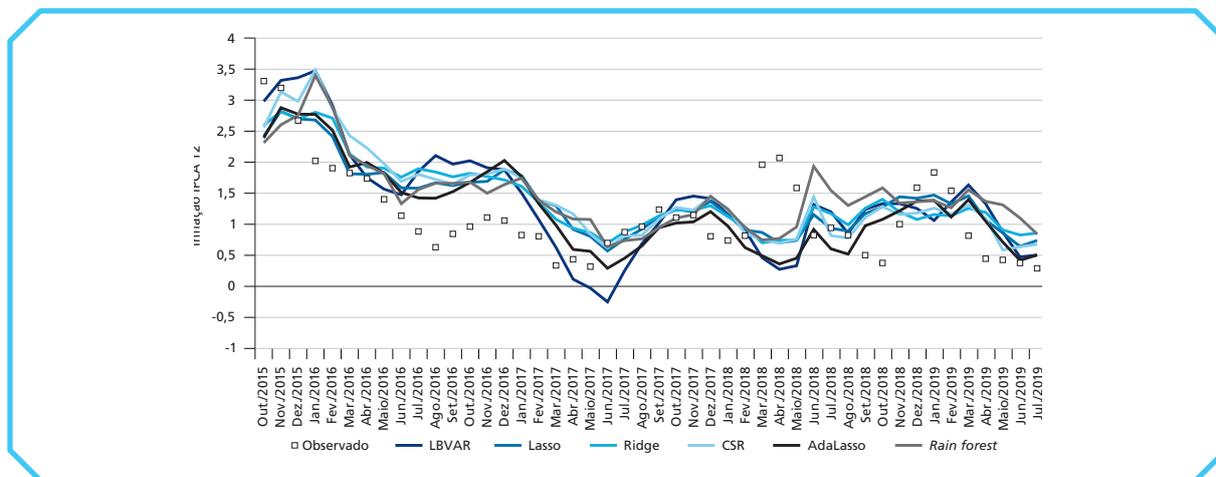
**GRÁFICO 1**

**IPCA: previsões um mês à frente e acumuladas de três, seis e doze meses à frente**

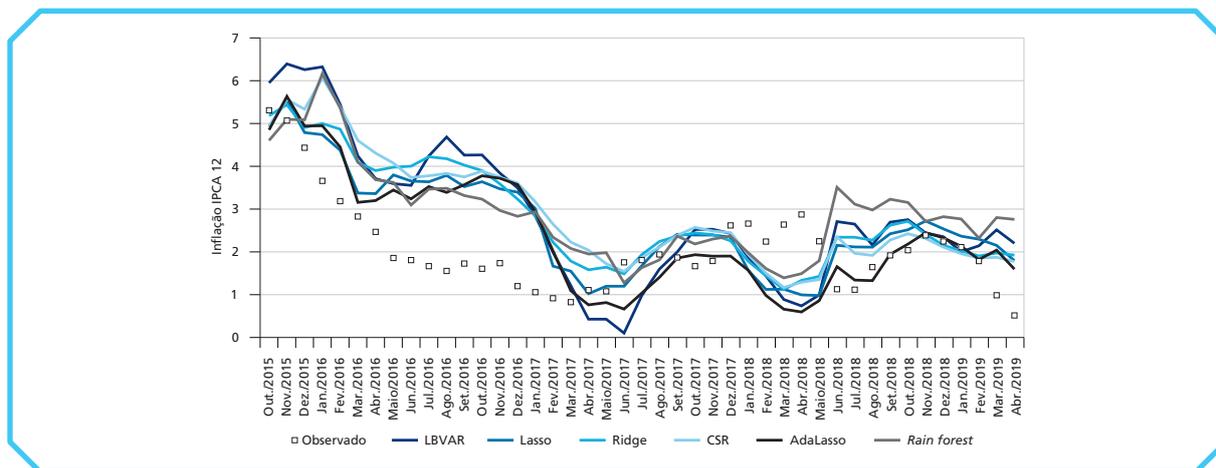
**1A – Um passo à frente**



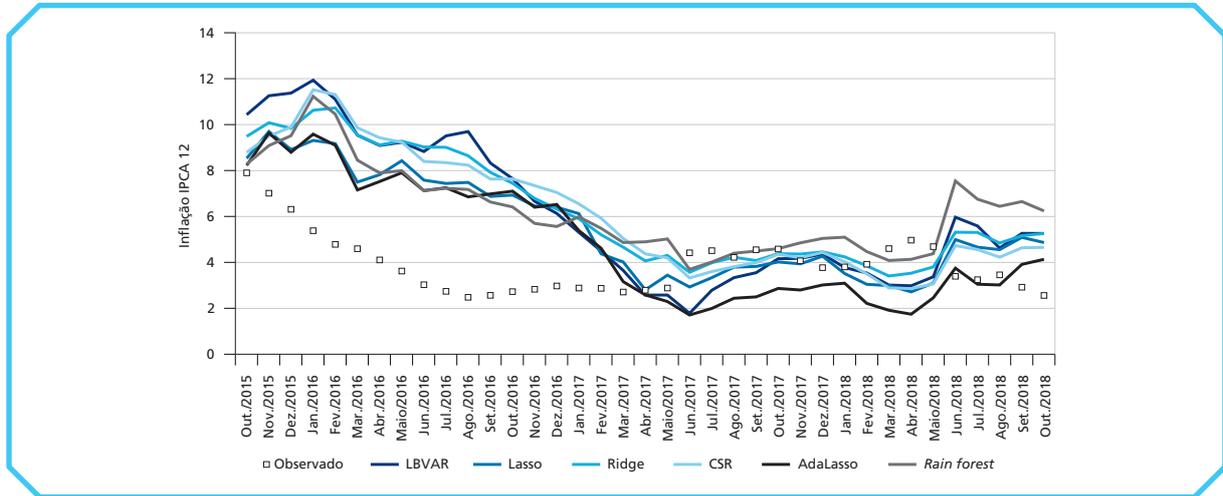
**1B – Três passos à frente**



**1C – Seis passos à frente**



1D – Doze passos à frente



Elaboração dos autores.

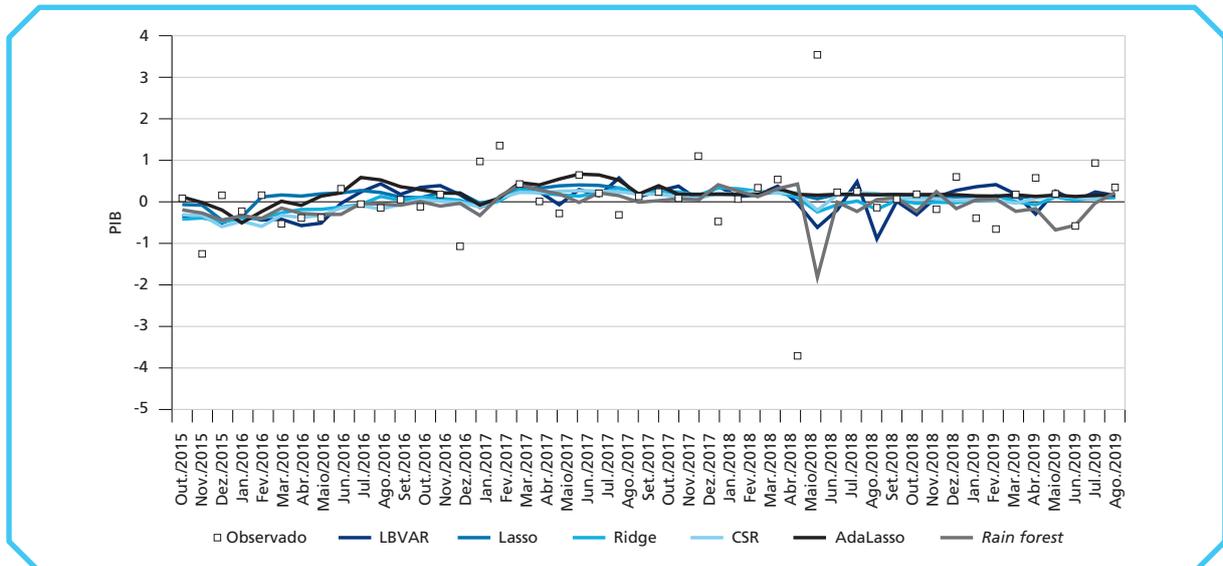
Obs.: LBVAR – *large Bayesian vector autoregression*.

O desempenho das previsões se agrava à medida que avançamos para 2016, o que coincide com a deterioração do cenário econômico e político brasileiro, melhorando a partir do segundo semestre de 2017. Note que há períodos em que a inflação observada está persistentemente abaixo das previsões.

**GRÁFICO 2**

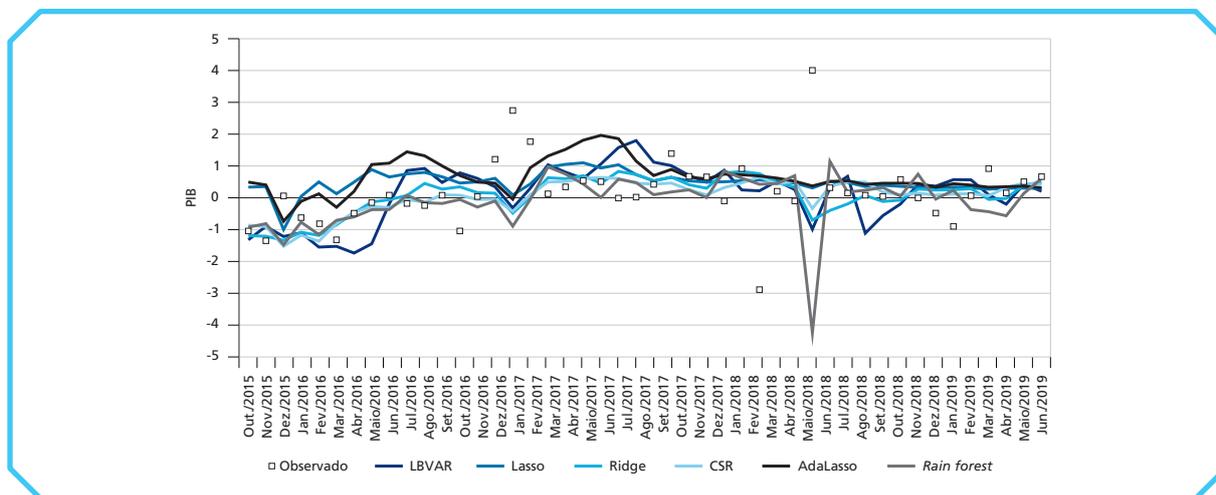
**PIB: previsões um mês à frente e acumuladas de três, seis e doze meses à frente**

2A – Um passo à frente

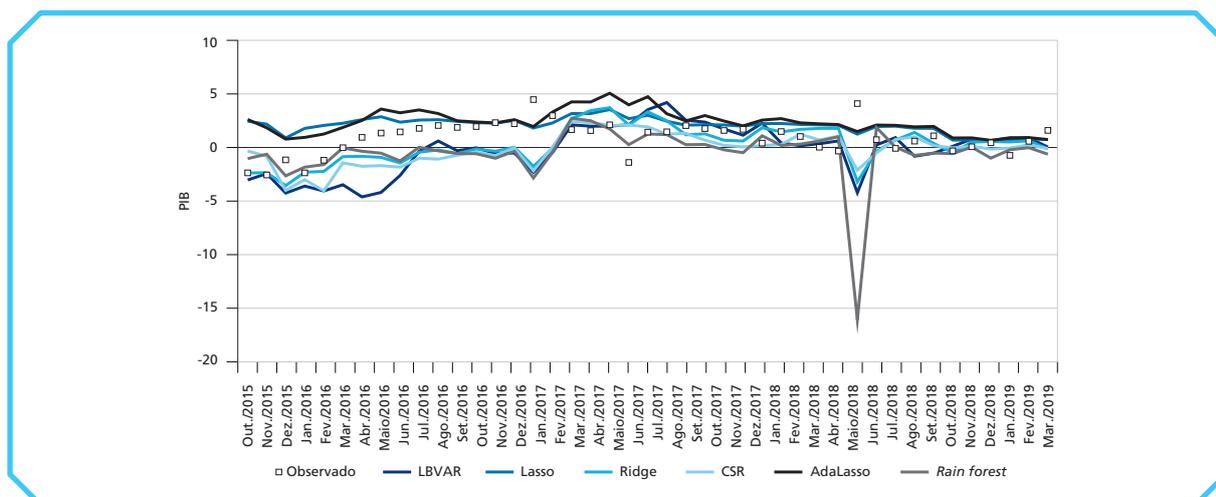


TEXTO para DISCUSSÃO

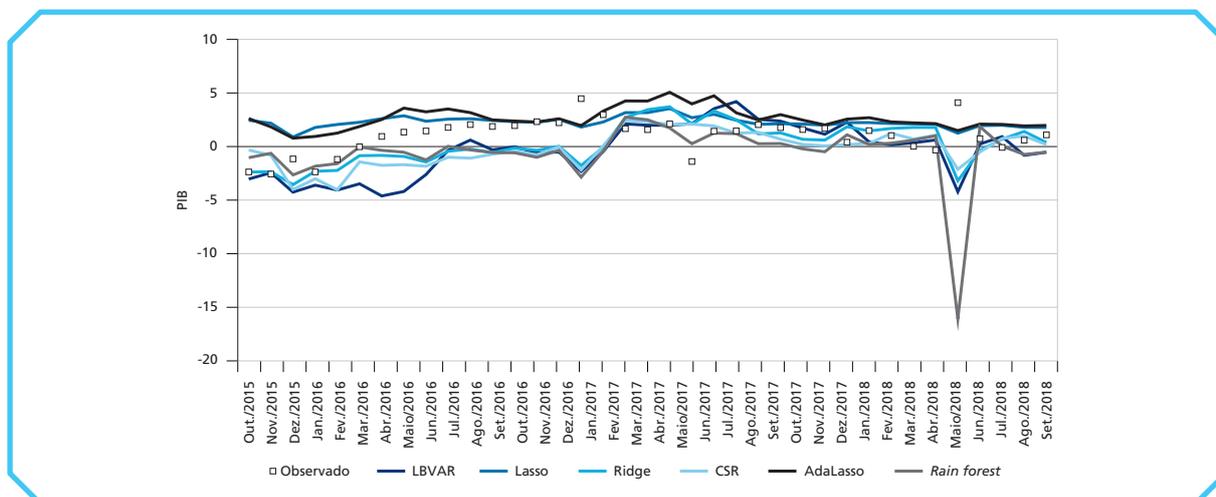
2B – Três passos à frente



2C – Seis passos à frente



2D – Doze passos à frente



Elaboração dos autores.

A partir do gráfico 2, notamos que, para horizontes curtos, todos os modelos se apresentam como uma versão mais suave do PIB observado, enquanto para horizontes mais longos podemos perceber que o *adaLasso* tende a ficar por cima dos demais e o *random forest*, a ficar por baixo. Vale ressaltar que o *random forest* reage exageradamente a mudanças no PIB, prevendo o oposto da realidade.

A fim de avaliar que variáveis são relevantes para previsão, identificamos aquelas selecionadas pelos métodos Lasso e *adaLasso* para prever o IPCA e o PIB, respectivamente, para os horizontes um e doze meses. Na previsão para o IPCA, chama a atenção o fato de que a natureza distinta dos preditores com maior participação muda de acordo com o horizonte de previsão. No curto prazo ( $T+1$ ), nota-se que a distribuição do poder explicativo, em muitas variáveis em que se destacam algumas medidas de inflação desagregada (monitorados e alimentos) – o Índice de Preços ao Consumidor (IPC), da FGV, a expectativa do *Focus* para o próximo mês e para os próximos três meses, a própria defasagem do IPCA, o total de provisões no mercado de crédito e uma variável de clima<sup>7</sup> –, aparece em mais de 50% das observações de teste.

No longo prazo, no entanto, além da manutenção do IPC/FGV e da expectativa do *Focus* para o próximo mês, a taxa de desemprego – Cadastro Geral de Empregados e Desempregados (Caged) e Pesquisa Nacional por Amostra de Domicílio (PNAD) – e as exportações aparecem na maior parte das janelas. A ausência de variáveis defasadas e de componentes de alto grau de variabilidade, como o clima, confirma que a inflação de longo prazo é explicada por movimentos mais perenes da economia.

Quanto à previsão de PIB, podemos concluir que o conjunto de variáveis selecionadas pelos modelos de regularização é muito restrito e há também uma segmentação por horizonte. No curto prazo ( $T+1$ ), aparecem o total de provisões no mercado de crédito e a expectativa de inflação para três meses. No longo prazo, a taxa de juros de duplicatas é a única variável selecionada de forma recorrente. Ainda que faça sentido a incorporação de variáveis do mercado de crédito, taxa de juros e de inflação, é importante destacar a ausência de variáveis típicas e indicadores coincidentes da atividade real (desemprego, vendas de veículos, comércio), o que merece investigação posterior mais detalhada.

---

7. Medida de temperatura dos oceanos fornecida pela National Oceanic and Atmospheric Administration (NOAA).

## 5 TESTES ADICIONAIS

As previsões dos modelos para o IPCA foram confrontadas a partir do teste de Diebold-Mariano, em que a comparação é feita por pares. Concluímos que, à medida que o horizonte aumenta, as previsões tendem a não se distinguir.

Vimos anteriormente que as previsões para o IPCA estão sobrestimadas desde 2016. As variáveis explicativas da base de dados foram escolhidas procurando maximizar a abrangência informacional. No entanto, muitas das variáveis deste conjunto são condicionadas pelas variáveis-alvo, gerando uma endogeneidade que talvez explique a incapacidade do modelo em antecipar quebras de regime. Para avaliar isso, fizemos uma seleção arbitrária<sup>8</sup> de variáveis que pretensamente são exógenas. Os modelos foram estimados para o IPCA e o PIB com o conjunto de dados completo e restrito e comparados por meio da estatística Theil-U.

Os resultados mostram que realizar previsões com um subconjunto de explicativas que são “exógenas” da economia brasileira não melhora o resultado, uma vez que o Theil-U não é melhor em nenhum dos casos, o que sugere que a seleção não melhorou a capacidade preditiva dos modelos.

## 6 CONCLUSÃO

Nossos resultados indicam que, para horizontes de previsão de curto prazo (um e três meses), os métodos de ML melhoram substancialmente as previsões de inflação em relação aos modelos tradicionais, incluindo a previsão de especialistas. Para o horizonte de um mês, o Lasso é o método com melhor desempenho, e a combinação dos melhores modelos de ML por horizonte é o destaque para o horizonte de três meses. No entanto, para horizontes de previsão mais longos (seis e doze meses), os modelos de ML e suas combinações não superam as expectativas de especialistas, ainda que as combinações de métodos de ML superem as metodologias tradicionais. Na previsão para o PIB, ocorre o contrário. Os modelos de ML não têm desempenho satisfatório para o horizonte mais curto, mas as combinações (três e doze meses) e o CSR (seis meses) superam os modelos tradicionais.

No geral, o desempenho da previsão é melhor no início da amostra. A partir de 2016, todos os métodos fornecem previsões que são geralmente acima dos valores de inflação observados. Como esse período coincide com uma perspectiva de deterioração do ambiente econômico e político brasileiro, vale a pena investigar a inclusão de novas variáveis e/ou métodos em

---

8. Além de expectativas e defasagens, incluem-se preço de *commodities*, juros externos, medidas de risco.

pesquisas futuras. Os testes que realizamos com seleção exógena de variáveis não mostraram resultados satisfatórios, e mesmo os modelos não lineares, que deveriam ter desempenho melhor, sobrestimam a inflação de forma consistente no período.

Em relação à seleção de variáveis, chama atenção a natureza distinta dos preditores de acordo com o horizonte de previsão. Na previsão de inflação, variáveis de estado da economia, como a taxa de desemprego (Caged e PNAD) e as exportações, aparecem no horizonte mais longo (doze meses). A ausência de variáveis mostra que o efeito autorregressivo e de componentes de alto grau de variabilidade, como o clima, confirmam que a inflação de longo prazo é explicada por fundamentos da economia. Quanto à previsão de PIB, podemos concluir que o conjunto de variáveis selecionadas pelos modelos de regularização é muito restrito e também há uma segmentação por horizonte.

No curto prazo ( $T+1$ ), o total de provisões no mercado de crédito e a expectativa de inflação para três meses são selecionados enquanto, no longo prazo, a taxa de juros de duplicatas é a única variável selecionada de forma recorrente. Ainda que faça sentido a incorporação de variáveis do mercado de crédito, taxa de juros e de inflação, é importante destacar a ausência de variáveis típicas e indicadores coincidentes da atividade real (desemprego, vendas de veículos, comércio), o que merece investigação posterior mais detalhada.

Em suma, os bons desempenhos obtidos com a aplicação dos métodos de ML para o Brasil sugerem que essa pode ser uma ferramenta promissora para a melhoria das previsões macroeconômicas no Brasil. Esta proposta de previsão não se restringe à variável inflação, podendo seu escopo se estender a outras variáveis macroeconômicas, como crédito, câmbio e atividade. Nesse sentido, um ponto de atenção é o tratamento de variáveis com forte componente sazonal, como a produção industrial. Testes preliminares indicam que as atuais rotinas não lidam bem com a sazonalidade quando somente tratada a partir da inclusão de variáveis *dummy*. Finalmente, o fraco desempenho preditivo em momentos de maior volatilidade traz à luz a necessidade de inclusão de novas variáveis que capturem melhor e de forma mais tempestiva os sinais da economia.

## REFERÊNCIAS

- BANBURA, M.; GIANNONE, D.; REICHLIN, L. Large Bayesian vector auto regressions. **Journal of Applied Econometrics**, v. 25, n. 1, p. 71-92, 2010.
- DOAN, T.; LITTERMAN, R.; SIMS, C. Forecasting and conditional projection using realistic prior distributions. **Econometric Reviews**, v. 3, n. 1, p. 1-100, 1984.
- ELLIOTT, G.; GARGANO, A.; TIMMERMANN, A. Complete subset regressions. **Journal of Econometrics**, v. 177, n. 2, p. 357-373, 2013.
- \_\_\_\_\_. Complete subset regressions with large-dimensional sets of predictors. **Journal of Economic Dynamics and Control**, v. 54(C), p. 86-110, 2015.
- GARCIA, M. G. P.; MEDEIROS, M. C.; VASCONCELOS, G. F. R. Real-time inflation forecasting with high-dimensional models: the case of Brazil, **International Journal of Forecasting**, v. 33, n. 3, p. 679-693, 2017.
- HANSEN, P. R.; LUNDE, A.; NASON, J. M. The model confidence set. **Econometrica**, v. 79, n. 2, p. 453-497, 2011.
- HOERL, A. E.; KENNARD, R. W.; BALDWIN, K. F. Ridge regression: some simulations. **Communications in Statistics**, v. 4, n. 2, p. 105-123, 1975.
- LITTERMAN, R. Forecasting with Bayesian vector autoregressions – five years of experience. **Journal of Business and Economic Statistics**, v. 4, p. 25-38, 1986.
- MEDEIROS, M. C. *et al.* Forecasting inflation in a data-rich environment: the benefits of machine Learning Methods. **Journal of Business and Economic Statistics**, v. 39, p. 1-45, 2018.
- MEDEIROS, M. C.; MENDES, E. F. Adaptive Lasso estimation for ARDL models with GARCH innovations. **Econometric Reviews**, v. 36, p. 622-637, 2017.
- TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. **Journal of the Royal Statistical Society**, v. 58, n. 1, p. 267-288, 1996. (Series B – Methodological).
- ZHAO, P.; YU, B. On model selection consistency of Lasso. **Journal of Machine Learning Research**, v. 7, p. 2541-2563, 2006.
- ZOU, H. The adaptive Lasso and its oracle properties. **Journal of the American Statistical Association**, v. 101, p. 1418-1429, 2006.
- ZOU, H.; HASTIE, T.; TIBSHIRANI, R. On the degrees of freedom of the lasso. **The Annals of Statistics**, v. 35, n. 5, p. 2173-2192, 2007.

# Ipea – Instituto de Pesquisa Econômica Aplicada

## EDITORIAL

### Coordenação

Aeromilson Trajano de Mesquita

### Assistentes da Coordenação

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

### Supervisão

Aline Cristine Torres da Silva Martins

### Revisão

Bruna Oliveira Ranquine da Rocha

Carlos Eduardo Gonçalves de Melo

Elaine Oliveira Couto

Lis Silva Hall

Luciana Bastos Dias

Marlon Magno Abreu de Carvalho

Rebeca Raimundo Cardoso dos Santos

Vivian Barros Volotão Santos

Débora Mello Lopes (estagiária)

Maria Eduarda Mendes Laguardia (estagiária)

### Editoração

Aline Cristine Torres da Silva Martins

Mayana Mendes de Mattos

Mayara Barros da Mota

### Capa

Aline Cristine Torres da Silva Martins

### Projeto Gráfico

Aline Cristine Torres da Silva Martins

*The manuscripts in languages other than Portuguese published herein have not been proofread.*

## **Missão do Ipea**

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.



**ipea** Instituto de Pesquisa  
Econômica Aplicada

MINISTÉRIO DA  
ECONOMIA



PÁTRIA AMADA  
**BRASIL**  
GOVERNO FEDERAL