

Previsão de Inflação Usando Modelos de Alta Dimensionalidade

Marco Aurélio Guerra

Coleta de Dados

Para previsão do IPCA mensal foram utilizados 70 variáveis¹, sendo estas retiradas de 3 fontes, Banco de dados do Banco Central (GetBCBdata, rbcB), Yahoo Finance (BatchGetSymbols) e Boletim de resultado do Tesouro Nacional². Para a realização da coleta de dados foi utilizado o código ‘Carrega_dados.R’, tendo como objeto de retorno “df.final” em data.frame.

```
source(Carrega_dados.R)

# Output df.final

# "tempo total coletando os dados : 8.31 mins"
```

Tratamento dos Dados

A variação e a expectativa do IPCA mensal, seguindo a metodologia utilizada em Figueiredo (2010) e Medeiros et al. (2019), foram transformadas em log retorno³. Para os dados coletados que possuíam frequência diária foi considerado os últimos dados observados no mês⁴.

```
### Exemplo de como foi feito a coleta do último dados observado do mês

df.ibov.diario_para_mensal <- df.ibov %>%
  tidyr::fill( 'price.adjusted' ) %>%
  group_by(strftime(ref.date, '%Y-%m')) %>%
  filter(ref.date == max(ref.date))
```

No caso das expectativas foi necessário desempilhar os dados para os diferentes horizontes de previsão.

```
### Exemplo de como foi desempilhado os dados

expectativas_desempilhado <- reshape(expectativas_empilhadas ,
  idvar = 'date', v.names = 'median',
  timevar = 't' , direction = 'wide')
```

Para evitar o problema de não estacionariedade, foi realizado um teste dick-fuller aumentado considerando a tendência e a deriva da série. Não havendo rejeição da hipótese nula de não estacionariedade⁵ os dados foram transformados em diferença dos logs⁶.

¹ descrição nas notas

² <https://www.tesourotransparente.gov.br/publicacoes/boletim-resultado-do-tesouro-nacional-rtn/2021/2>. Utilizei a planilha, pois os dados API não é atualizado desde de agosto de 2020.

³ $\pi_t = \log\left(\frac{IPCA_t}{IPCA_{t-1}} + 1\right)$

⁴ ‘taxa de câmbio Euro venda’, ‘Saldo diário de depósitos de poupança’, ‘Captação líquida diária de depósitos de poupança’, ‘Taxa média flutuantes DI de depósitos a prazo (CDB/RDB)’, ‘Taxa de juros - Meta Selic definida pelo Copom’, ‘índice bovespa’

⁵ Modelos testados tinham a especificação de 1 a 4 lags da variável

⁶ $x_t = \log(X_t) - \log(X_{t-1})$

```
# teste de estacionariedade

teste <- aTSA::adf.test(as.ts(dados.transformados[,coluna]) ,output = FALSE)

# transformando as variaveis não estacionarias
if (any(teste$type3[,3] > .01) ){

  print(paste0('a serie ', colnames(dados.transformados)[coluna], ' NÃO é estacionária'))

  dados.transformados[,coluna] <- diff(log(dados.transformados[,coluna]))

} else {

  print(paste0('a serie ', colnames(dados.transformados)[coluna], ' é estacionária'))

}
```

Modelagem

A modelagem utilizada no presente trabalho se baseia em dois artigos. O artigo escrito por Garcia, Medeiros, and Vasconcelos (2017) aplica modelos de alta dimensionalidade para previsão de inflação no Brasil. O trabalho de Medeiros et al. (2019) amplia o número de modelos utilizados no texto de Garcia, Medeiros, and Vasconcelos (2017) e os aplica para estimação do índice de preços da economia Norte Americana. Ambos os artigos disponibilizaram os códigos utilizados na forma de um pacote para o R⁷⁸.

As previsões do IPCA mensal são estimações *out of sample* para 108 meses, onde o modelo é reestimado a cada periodo (*rolling window*) com uma janela de 106 meses, sendo observado o período de 2003-01 até 2020-12. Em todos os modelos foram realizados as estimações do IPCA para 13 horizontes, $h = \{1, 2, \dots, 13\}$. As estimativas dos modelos são avaliadas pela raiz quadrática média dos erros (RMSE)⁹ e pelo o Erro Médio Absoluto (MAE)¹⁰.

Para estimação da variação mensal do IPCA foram utilizados 14 modelos e a média de todas as projeções. A lista de modelos a baixo foi baseada no artigo de Medeiros et al. (2019):

- Modelos para benchmark : AR, Random Walk (RW), Componente Não Observado de Volatilidade Estocástica (UCSV)
- Modelos de Shrinkage : Lasso , adaLasso, flex-adaLasso
- Modelos de fatores : Modelo de Fatores, Fatores Alvo, Boosting de Fatores
- Modelos de Ensemble : Regressões de subconjunto completas (CSR)
- Modelos Random Forest : Random Forest (RF), Random Forest com Mínimos Quadrados Ordinários, Random Forest com AdaLasso
- Modelo Vetor Auto Regressivo : Var Bayesiano
- Média das previsões de todos os modelos anteriores

⁷Ver <https://github.com/gabrielrvsc/ForecastingInflation> e <https://github.com/gabrielrvsc/HDeconometrics>

⁸Utilizei os scripts de funções individualmente. Além disso, para conseguir rodar as funções foram necessárias algumas mudanças nos códigos, pois encontrei alguns bugs.

$$^9 RMSE(\hat{Y}) = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}$$

$$^{10} MAE(\hat{Y}) = \frac{\sum_{t=1}^T |\hat{y}_t - y_t|}{T}$$

TODO : mudar isso aqui

"tempo total rodando os modelos : 8.01 hours"

Resultados

Os resultados de todos os modelos são guardados em um objeto do tipo lista ('todas_previsoes_por_horizonte'). Com essas estimativas foram calculados o RMSE e o MAE para todos os modelos e horizontes. Para auxiliar na avaliação dos resultados do RMSE e MAE, o modelo de random walk foi utilizado como referência através seguinte calculo :

$$ScoreRelativo = \frac{RMSE(\hat{Y}_{modelo})}{RMSE(\hat{Y}_{RandomWalk})}$$

A tabela a seguir contém os resultados do Score Relativo calculado para o RMSE e o MAE. O resultado dessa análise é de que o modelo mais preciso, de maneira geral, é a média de todos os modelos. No entanto, para o primeiro período os modelos Lasso e o Random Forest com Mínimos Quadrados Ordinários possuem um resultado melhor do RMSE

modelo	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13
benchmark													
ar	0.92 (0.92)	0.84 (0.84)	0.80 (0.80)	0.76 (0.76)	0.77 (0.77)	0.76 (0.76)	0.82 (0.82)	0.83 (0.83)	0.93 (0.93)	0.96 (0.96)	0.92 (0.92)	0.91 (0.91)	0.85 (0.85)
rw	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
ucsv	1.00 (1.00)	0.96 (0.96)	0.99 (0.99)	0.96 (0.96)	0.96 (0.96)	0.93 (0.93)	0.95 (0.95)	0.93 (0.93)	0.93 (0.93)	0.98 (0.98)	0.96 (0.96)	0.95 (0.95)	0.96 (0.96)
modelos													
lasso	0.69 (0.69)	0.76 (0.76)	0.74 (0.74)	0.70 (0.70)	0.72 (0.72)	0.74 (0.74)	0.80 (0.80)	0.78 (0.78)	0.90 (0.90)	0.86 (0.86)	0.85 (0.85)	0.97 (0.97)	0.88 (0.88)
ridge	1.11 (1.11)	0.87 (0.87)	0.83 (0.83)	0.75 (0.75)	0.75 (0.75)	0.73 (0.73)	0.79 (0.79)	0.80 (0.80)	0.90 (0.90)	0.93 (0.93)	0.89 (0.89)	0.89 (0.89)	0.83 (0.83)
adalasso	0.69 (0.69)	0.76 (0.76)	0.74 (0.74)	0.70 (0.70)	0.72 (0.72)	0.74 (0.74)	0.80 (0.80)	0.78 (0.78)	0.90 (0.90)	0.86 (0.86)	0.85 (0.85)	0.97 (0.97)	0.88 (0.88)
flex_adalasso	0.69 (0.69)	0.76 (0.76)	0.74 (0.74)	0.70 (0.70)	0.72 (0.72)	0.74 (0.74)	0.80 (0.80)	0.78 (0.78)	0.90 (0.90)	0.86 (0.86)	0.85 (0.85)	0.97 (0.97)	0.88 (0.88)
fact	0.96 (0.96)	0.84 (0.84)	0.82 (0.82)	0.79 (0.79)	0.78 (0.78)	0.73 (0.73)	0.75 (0.75)	0.78 (0.78)	0.83 (0.83)	0.92 (0.92)	0.83 (0.83)	0.82 (0.82)	0.79 (0.79)
target_fact	0.88 (0.88)	0.79 (0.79)	0.79 (0.79)	0.76 (0.76)	0.75 (0.75)	0.71 (0.71)	0.81 (0.81)	0.80 (0.80)	0.91 (0.91)	0.91 (0.91)	0.86 (0.86)	0.84 (0.84)	0.79 (0.79)
boosting_fact	1.02 (1.02)	0.83 (0.83)	0.81 (0.81)	0.76 (0.76)	0.70 (0.70)	0.68 (0.68)	0.74 (0.74)	0.70 (0.70)	0.77 (0.77)	0.79 (0.79)	0.77 (0.77)	0.78 (0.78)	0.77 (0.77)
csr	0.75 (0.75)	0.74 (0.74)	0.73 (0.73)	0.68 (0.68)	0.67 (0.67)	0.68 (0.68)	0.74 (0.74)	0.72 (0.72)	0.80 (0.80)	0.85 (0.85)	0.84 (0.84)	0.82 (0.82)	0.75 (0.75)
rf	0.78 (0.78)	0.75 (0.75)	0.72 (0.72)	0.65 (0.65)	0.66 (0.66)	0.64 (0.64)	0.71 (0.71)	0.71 (0.71)	0.78 (0.78)	0.80 (0.80)	0.78 (0.78)	0.79 (0.79)	0.76 (0.76)
rfols	0.70 (0.70)	0.74 (0.74)	0.72 (0.72)	0.65 (0.65)	0.64 (0.64)	0.64 (0.64)	0.70 (0.70)	0.71 (0.71)	0.81 (0.81)	0.83 (0.83)	0.81 (0.81)	0.83 (0.83)	0.80 (0.80)
rf_adalasso	0.80 (0.80)	0.81 (0.81)	0.74 (0.74)	0.69 (0.69)	0.71 (0.71)	0.70 (0.70)	0.79 (0.79)	0.81 (0.81)	0.82 (0.82)	0.82 (0.82)	0.80 (0.80)	0.83 (0.83)	0.80 (0.80)
lbvar	0.82 (0.82)	0.74 (0.74)	0.74 (0.74)	0.69 (0.69)	0.71 (0.71)	0.71 (0.71)	0.78 (0.78)	0.80 (0.80)	0.90 (0.90)	0.94 (0.94)	0.91 (0.91)	0.91 (0.91)	0.85 (0.85)
ucsv	1.00 (1.00)	0.96 (0.96)	0.98 (0.98)	0.96 (0.96)	0.96 (0.96)	0.93 (0.93)	0.95 (0.95)	0.93 (0.93)	0.93 (0.93)	0.98 (0.98)	0.96 (0.96)	0.95 (0.95)	0.96 (0.96)

modelo	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13
E(modelos)	0.72 (0.72)	0.70 (0.70)	0.68 (0.68)	0.64 (0.64)	0.65 (0.65)	0.61 (0.61)	0.68 (0.68)	0.68 (0.68)	0.74 (0.74)	0.76 (0.76)	0.73 (0.73)	0.76 (0.76)	0.72 (0.72)
FOCUS	0.99 (0.99)	0.77 (0.77)	0.75 (0.75)	0.67 (0.67)	0.68 (0.68)	0.67 (0.67)	0.73 (0.73)	0.75 (0.75)	0.84 (0.84)	0.87 (0.87)	0.83 (0.83)	0.83 (0.83)	0.77 (0.77)
TOP5FOCUS	1.01 (1.01)	0.78 (0.78)	0.78 (0.78)	0.68 (0.68)	0.68 (0.68)	0.68 (0.68)	0.75 (0.75)	0.77 (0.77)	0.86 (0.86)	0.88 (0.88)	0.83 (0.83)	0.83 (0.83)	0.77 (0.77)

Nota: Os números em parentesis são os valores do Erro Médio Absoluto (MAE) o restante são as raízes quadráticas médias dos erros (RMSE)

Bibliografia

Figueiredo, Francisco. 2010. “Forecasting Brazilian Inflation Using a Large Data Set.” *Working Paper Series - BACEN* 228 (December): 1–56. <https://www.bcb.gov.br/pec/wps/ingl/wps228.pdf>.

Garcia, Medeiros, and Vasconcelos. 2017. “Real-Time Inflation Forecasting with High-Dimensional Models, the Case of Brazil.” *International Journal of Forecasting* 33 (4): 679–93. <https://doi.org/10.1016/j.ijforecast.2017.02.002>.

Medeiros, Vasconcelos, Veiga, and Zilberman. 2019. “Forecasting Inflation in a Data-Rich Environment, the Benefits of Machine Learning Methods.” *Journal of Business & Economic Statistics* 39 (1): 98–119. <https://doi.org/10.1080/07350015.2019.1637745>.

Notas

Código/font	Nome da Variável	Unidade	Tipo
433 - BACEN	IPCA	Var. % mensal	Preço e dinheiro
189 - BACEN	IGP-M	Var. % mensal	Preço e dinheiro
190 - BACEN	IGP-DI	Var. % mensal	Preço e dinheiro
7447 - BACEN	IGP-10	Var. % mensal	Preço e dinheiro
1788 - BACEN	Base monetária restrita (saldo em final de período)	u.m.c (mil)	Preço e dinheiro
27791 - BACEN	M1 (saldo em final de período) - Novo	u.m.c (mil)	Preço e dinheiro
27810 - BACEN	M2 (saldo em final de período) - Novo	u.m.c (mil)	Preço e dinheiro
27813 - BACEN	M3 (saldo em final de período) - Novo	u.m.c (mil)	Preço e dinheiro
27815 - BACEN	M4 (saldo em final de período) - Novo	u.m.c (mil)	Preço e dinheiro
24369 - BACEN	Taxa de desocupação - PNADC	porcentagem	Emprego
24348 - BACEN	Número de horas trabalhadas – indústria de transformação (2006=100)	índice	Emprego
28766 - BACEN	Estoque de empregos formais – Indústrias de transformação	unidades	Emprego

Código/font	Nome da Variável	Unidade	Tipo
3695 - BACEN	Taxa de câmbio - Livre - Dólar americano (compra) - Fim de período - mensal	R\$ u.m.c	cambio e finanças
21619 - BACEN	Taxa de câmbio - Livre - Euro (venda)	R\$ u.m.c	cambio e finanças
Yahoo Finance	Ibov	Pontos	cambio e finanças
23 - BACEN	Saldo diário de depósitos de poupança - SBPE e rural	u.m.c (mil)	cambio e finanças
24 - BACEN	Captação líquida diária de depósitos de poupança - SBPE e rural	u.m.c (mil)	cambio e finanças
256 - BACEN	Taxa de juros de longo prazo - TJLP	per. a.a.	cambio e finanças
432 - BACEN	Taxa de juros - Meta Selic definida pelo Copom	per. a.a.	cambio e finanças
1157 - BACEN	Taxa média flutuantes DI de depósitos a prazo (CDB/RDB) - Total	per. a.d.	cambio e finanças
22701 - BACEN	Transações correntes - mensal - saldo	US\$ (milhões)	Governo e transações internacionais
22707 - BACEN	Balança comercial - Balanço de Pagamentos - mensal - saldo	US\$ (milhões)	Governo e transações internacionais
4503 - BACEN	Dívida Pública Líquida como Porcentagem do PIB	Porcentagem	Governo e transações internacionais
10825 - BACEN	Dívida Pública Líquida	US\$ (milhões)	Governo e transações internacionais
4480 - BACEN	Dívida Interna do Governo Central	US\$ (milhões)	Governo e transações internacionais
4479 - BACEN	Dívida Interna Líquida do Governo Central e Banco Central	US\$ (milhões)	Governo e transações internacionais
4472 - BACEN	Dívida Líquida dos Estados	US\$ (milhões)	Governo e transações internacionais
4494 - BACEN	Dívida Líquida dos Estados com estrangeiros	US\$ (milhões)	Governo e transações internacionais
4473 - BACEN	Dívida Líquida dos Municípios	US\$ (milhões)	Governo e transações internacionais
4495 - BACEN	Dívida Líquida dos Municípios com Estrangeiros	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Receita PIS & PASEP	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Revenue Receita do governo central	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Gasto total do Governo Central	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Renda de Imposto sobre Importação	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Renda do Imposto Sobre produto industrializados	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Receita de outros impostos	US\$ (milhões)	Governo e transações internacionais
Planilha do Tesouro	Receita do Sistema de segurança social	US\$ (milhões)	Governo e transações internacionais
BACEN	Mediana das Expectativas de mercado para IPCA em t (para t de 1 à 13)	Var. % mensal do Índice	Expectativas

Código/font	Nome da Variável	Unidade	Tipo
BACEN	desvio padrão das Expectativas de mercado (para t igual a 1, 2 e 12)	NA	Expectativas
BACEN	Mediana das Expectativas TOP5 de mercado para IPCA em t (para t de 1 à 13)	Var. % mensal do Índice	Expectativas