Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models

Gustavo Silva Araujo (Banco Central do Brasil)

Wagner Piazza Gaglianone (Banco Central do Brasil)

CEMLA – XXV Meeting of the Central Bank Researchers Network

28/10/2020

Introduction

- Machine Learning (ML) is often described as *the art and science of pattern recognition*;
- Data-driven approach, mild assumptions about the data;
- Encompasses a wide variety of models;
- Varian (2014): *"...the growing amounts of data and ever complex-growing relationships warrant the usage of machine learning approaches in economics"*
- **Objective of this paper**: Forecast the monthly inflation (IPCA) in Brazil using a large number of variables and a diverse set of methods.

Machine Learning methods

- We use the following ML (supervised) approaches:

  - *Elastic Net, Ridge Regression, Lasso, Adaptive Lasso;*

  - *Factor Models*;

  - *Random Forest, Quantile Regression Forest*.

- The first 4 methods are regularization techniques that introduce penalties for overfitting the data.

- The last 2 methods are nonparametric approaches, based on the recursive binary partitioning of the covariate space, which can deal with very large number of explanatory variables.

ML estimation and forecasting

Our goal is to forecast the inflation rate y_{t+h} using a set of predictors $\widetilde{x}_t$ (*direct forecasts*), as follows:

$$y_{t+h} = \Phi_h\left(\widetilde{x}_t\right) + \varepsilon_{t+h}, \tag{1}$$

where $\Phi_h(\cdot)$ is a nonlinear function, $x_t = \left(x_{1,t}, \ldots, x_{n,t}\right)$ is a set of n predictors, $\widetilde{x}_t = \left(x_t, x_{t-1}, \ldots, x_{t-s}, c, d_{i,t}\right)'$, c is the intercept, $d_{i,t}$ are seasonal dummies.

In some cases (e.g., elastic net), the mapping $\Phi_h(\cdot)$ is linear:

$$y_{t+h} = \widetilde{x}_t'\beta_h + \varepsilon_{t+h}. \tag{2}$$

Elastic Net

The elastic net is a regularization (and variable selection method) proposed by Zou and Hastie (2005), as a generalization of *LASSO*.

For a nonnegative shrinkage parameter λ, and a combination parameter $\alpha \in [0, 1]$, the elastic net solves:

$$\widehat{\beta} = \underset{\{\beta_1,\ldots,\beta_k\}}{\arg\min} \left(\frac{1}{T} \sum_{t=1}^{T} \left(y_{t+h} - \sum_{j=1}^{k} x'_{j,t} \beta_j \right)^2 + \lambda P_\alpha(\beta) \right), \tag{3}$$

where

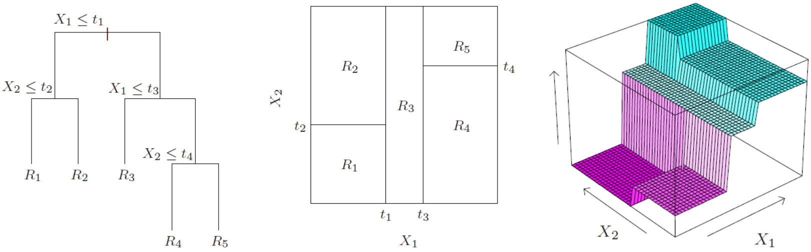
$$P_\alpha(\beta) = \sum_{j=1}^{k} \alpha \left| \beta_j \right| + \frac{(1-\alpha)}{2} \beta_j^2. \tag{4}$$

Ridge: $\alpha = 0$; LASSO: $\alpha = 1$; Adaptive LASSO: two-step approach (adaptive weights for penalizing different coefficients).

Random Forest

- Introduced as a machine learning tool in Breiman (2001).
- Very popular and powerful tool for high-dimensional regression and classification.
- Collection of regression trees, designed to reduce the prediction variance by using bootstrap aggregation (*bagging*).
- A *regression tree* is a nonparametric model based on the recursive binary partitioning of the covariate space X.

Figure 1 - Example of a recursive binary splitting in a regression tree



Source: Hastie et al. (2009).

Quantile Regression Forest

- RF approximates the conditional mean of Y by constructing a weighted average over the sample observations of Y.
- Conditional quantiles can be inferred with quantile regression forests (QRF); a generalization of RF proposed by Meinshausen (2006).
- Non-parametric way of estimating conditional quantiles for a high-dimensional set of predictors.

Factor Models

Factor model 1 (direct forecast): Let $x_{i,t}$ be the observed data and consider the factor representation:

$$x_{i,t} = \lambda_i' F_t + e_{i,t}, \tag{5}$$

where F_t is a vector of common factors and λ_i is a vector of loadings. Direct forecast approach:

$$y_{t+h} = \beta_h F_t + \varepsilon_{t+h}. \tag{6}$$

Factor model 2 (iterated forecast): inflation and factors are contemporaneous:

$$y_t = \gamma F_t + \varepsilon_t, \tag{7}$$

and F_t follows a VAR process (Bańbura et al., 2013).

Factor Models

Factor models 3 and 4 (with targeted predictors):

- The same factor models, but now using a subset of predictors.
- Bai and Ng (2008): the forecast accuracy can be improved by selecting the predictors.
- The core idea is that irrelevant predictors employed to build a factor model only add noise into the analysis, and thus produce factors with a poor predictive performance.

Traditional econometric models

RW: Random walk model, such that $E\left(y_{t+h} - y_t \mid \mathcal{F}_t\right) = 0$.

RW-AO: Variant considered by Atkeson and Ohanian (2001), assuming the average inflation over the previous 4 years as the forecast for y_{t+h}.

ARMA: One of the most common statistical models used for time-series forecasting. The best model in our exercise is the AR(1), according to the Schwarz information criterion.

Traditional econometric models

VAR: 1 lag and 4 endogenous variables: IPCA (market prices), IPCA (regulated prices), $\Delta \ln(\text{M4})$, and $\Delta \ln(\text{FX-rate})$.

PC-backward: Phillips Curve (PC) with past inflation, imported inflation and output gap. PC for market inflation and ARMA for the regulated and monitored prices inflation.

PC-hybrid: New Keynesian version of the PC, with backward and forward looking terms, imported inflation and output gap. Again, PC for market inflation and ARMA for regulated and monitored prices inflation.

Disaggregate Forecasts

Bottom-up approach:

- Main idea is to exploit different dynamics estimated for different IPCA sub-indexes:

  (i) regulated and monitored prices
  (ii) tradables
  (iii) non-tradables

- Methods: ARFIMA, Adalasso, Random Forest.

- Aggregate the individual forecasts using the IPCA weights

  (computed via rolling window)

Forecast Combination

Elliott et al. (2015): *"By diversifying across multiple models, combinations typically deliver more stable forecasts than those associated with individual models."*

Here, we employ 8 forecast combination methods:

- Median, Mean, Mean (selected methods),
- Adalasso, Random Forest,
- Complete Subset Regression (CSR): combine forecasts from all possible linear regression models that keep the number of predictors fixed,
- Granger and Ramanathan (1984): OLS regression of IPCA onto individual forecasts,
- Constrained Least Squares (CLS): same OLS above, with additional constraints: no intercept, $\beta_i \geq 0$, $\sum \beta_i = 1$.

Empirical Exercise - Data

- Database: 165 variables (Thomson Reuters, EPU, SGS, Anbima, Inmet,...)
- Monthly data: January 2004 - June 2019
- Number of predictors: 507 (=165*3+12 seasonal dummies)
- Series are first-differenced when necessary (KPSS test)
- Forecast horizon from $h = 1, ..., 12$ months
- Sample divided in three parts: 5 years (training sample) + 4 years (forec.comb. weights) + 6.5 years (out-of-sample evaluation)
- **36 forecasting methods** = 6 traditional + 10 ML + 3 disaggregate forec. + 13 forec.comb. + 2 BEI (breakeven inflation) + 2 Focus (survey)

Empirical Exercise - Results

Table 1 - Mean Squared Error (MSE)

Model	Name	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12
1	RW	0.093	0.152	0.160	0.185	0.223	0.211	0.219	0.210	0.170	0.161	0.176	0.191
2	RW-AO	0.121	0.125	0.127	0.127	0.129	0.132	0.135	0.139	0.139	0.143	0.145	0.145
3	AR	0.076	0.106	0.109	0.115	0.122	0.122	0.124	0.125	0.124	0.126	0.129	0.128
4	VAR	0.075	0.110	0.111	0.112	0.125	0.123	0.122 *	0.123	0.122	0.124	0.128	0.129
5	PC backward	0.075	0.111	0.121	0.125	0.157	0.160	0.160	0.145	0.124	0.121	0.125	0.120
6	PC hybrid	0.072 *	0.100	0.119	0.132	0.169	0.179	0.186	0.163	0.139	0.130	0.128	0.117
7	Factor model1	0.069	0.086 *	0.105	0.108	0.123	0.134	0.126	0.123	0.124	0.119	0.136	0.134
8	Factor model2	0.067	0.087 *	0.102	0.112	0.116	0.116	0.113	0.115	0.116	0.121	0.125	0.125
9	Factor model3	0.063 **	0.080 *	0.093	0.102	0.103	0.126	0.143	0.126	0.122	0.123	0.132	0.139
10	Factor model4	0.073	0.093	0.091	0.100	0.110	0.111	0.113	0.114	0.117	0.123	0.126	0.129
11	Elastic net	0.079	0.106	0.100	0.102	0.098 **	0.101 **	0.122	0.129	0.146	0.130	0.132	0.128
12	Lasso	0.078	0.101	0.099	0.097	0.100 **	0.101 *	0.127	0.136	0.147	0.139	0.132	0.130
13	Adalasso	0.077	0.098	0.099	0.103	0.103 *	0.102 **	0.141	0.135	0.126	0.143	0.141	0.154
14	Ridge regression	0.073	0.092 *	0.093	0.094 *	0.102 *	0.110	0.113	0.121	0.120	0.128	0.132	0.126
15	Random forest	0.073 **	0.089 **	0.092 **	0.097 **	0.104 **	0.111	0.115	0.117	0.117	0.119	0.121	0.119
16	Quant.reg.forest	0.072 ***	0.089 **	0.092 **	0.097 **	0.103 **	0.109	0.115	0.117	0.116	0.117	0.120	0.118
17	Disag. ARFIMA	0.081	0.108	0.114	0.117	0.123	0.122	0.125	0.126	0.124	0.127	0.131	0.133
18	Disag. Adalasso	0.091	0.096	0.101	0.102	0.108	0.107	0.109	0.108	0.118	0.122	0.123	0.129
19	Disag. RF	0.072 *	0.091 *	0.094 *	0.097 *	0.104 *	0.110	0.115	0.120	0.117	0.120	0.121	0.122
20	F.comb.Median	0.070 **	0.093 **	0.095 *	0.098 **	0.106 **	0.109 **	0.116	0.120	0.117	0.119	0.122	0.123
21	F.comb.Mean	0.068 **	0.088 **	0.093 *	0.097 **	0.107 **	0.109 **	0.116	0.116	0.113	0.113	0.120	0.123
22	F.comb.Mean2	0.068 **	0.088 **	0.092 *	0.093 **	0.102 **	0.106 **	0.118	0.118	0.114	0.116	0.122	0.127
23	F.comb.GR	0.067 *	0.097	0.113	0.127	0.136	0.198	0.170	0.164	0.203	0.209	0.216	0.215
24	F.comb.CLS	0.063 **	0.086 **	0.091	0.093	0.099	0.120	0.119	0.119	0.118	0.120	0.129	0.127
25	F.comb.CSR	0.070	0.095	0.100	0.103	0.110	0.126	0.135	0.129	0.141	0.151	0.168	0.183
26	F.comb.Adalasso	0.069	0.091	0.108	0.119	0.116	0.145	0.169	0.170	0.193	0.203	0.222	0.195
27	F.comb.RF	0.075	0.091	0.101	0.110	0.133	0.135	0.157	0.144	0.173	0.161	0.202	0.183
	Best model	***24***	***9***	***10***	***22***	***11***	***11***	***18***	***18***	***21***	***21***	***16***	***6***

Table 2 - Mean Squared Error (MSE)

Model	Name	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12
1	RW	0.093	0.152	0.160	0.185	0.223	0.211	0.219	0.210	0.170	0.161	0.176	0.191
2	RW-AO	0.121	0.125	0.127	0.127	0.129	0.132	0.135	0.139	0.139	0.143	0.145	0.145
3	AR	0.076	0.106	0.109	0.115	0.122	0.122	0.124	0.125	0.124	0.126	0.129	0.128
4	VAR	0.075	0.110	0.111	0.112	0.125	0.123	0.122 *	0.123	0.122	0.124	0.128	0.129
5	PC backward	0.075	0.111	0.121	0.125	0.157	0.160	0.160	0.145	0.124	0.121	0.125	0.120
6	PC hybrid	0.072 *	0.100	0.119	0.132	0.169	0.179	0.186	0.163	0.139	0.130	0.128	0.117
7	Factor model1	0.069	0.086 *	0.105	0.108	0.123	0.134	0.126	0.123	0.124	0.119	0.136	0.134
8	Factor model2	0.067	0.087 *	0.102	0.112	0.116	0.116	0.113	0.115	0.116	0.121	0.125	0.125
9	Factor model3	0.063 **	0.080 *	0.093	0.102	0.103	0.126	0.143	0.126	0.122	0.123	0.132	0.139
10	Factor model4	0.073	0.093	0.091	0.100	0.110	0.111	0.113	0.114	0.117	0.123	0.126	0.129
11	Elastic net	0.079	0.106	0.100	0.102	0.098 **	0.101 **	0.122	0.129	0.146	0.130	0.132	0.128
12	Lasso	0.078	0.101	0.099	0.097	0.100 **	0.101 *	0.127	0.136	0.147	0.139	0.132	0.130
13	Adalasso	0.077	0.098	0.099	0.103	0.103 *	0.102 **	0.141	0.135	0.126	0.143	0.141	0.154
14	Ridge regression	0.073	0.092 *	0.093	0.094 *	0.102 *	0.110	0.113	0.121	0.120	0.128	0.132	0.126
15	Random forest	0.073 **	0.089 **	0.092 **	0.097 **	0.104 **	0.111	0.115	0.117	0.117	0.119	0.121	0.119
16	Quant.reg.forest	0.072 ***	0.089 **	0.092 **	0.097 **	0.103 **	0.109	0.115	0.117	0.116	0.117	0.120	0.118
17	Disag. ARFIMA	0.081	0.108	0.114	0.117	0.123	0.122	0.125	0.126	0.124	0.127	0.131	0.133
18	Disag. Adalasso	0.091	0.096	0.101	0.102	0.108	0.107	0.109	0.108	0.118	0.122	0.123	0.129
19	Disag. RF	0.072 *	0.091 *	0.094 *	0.097 *	0.104 *	0.110	0.115	0.120	0.117	0.120	0.121	0.122
20	F.comb.Median	0.070 **	0.093 **	0.095 *	0.098 **	0.106 **	0.109 **	0.116	0.120	0.117	0.119	0.122	0.123
21	F.comb.Mean	0.068 **	0.088 **	0.093 *	0.097 **	0.107 **	0.109 **	0.116	0.116	0.113	0.113	0.120	0.123
22	F.comb.Mean2	0.068 **	0.088 **	0.092 *	0.093 **	0.102 **	0.106 **	0.118	0.118	0.114	0.116	0.122	0.127
23	F.comb.GR	0.067 *	0.097	0.113	0.127	0.136	0.198	0.170	0.164	0.203	0.209	0.216	0.215
24	F.comb.CLS	0.063 **	0.086 **	0.091	0.093	0.099	0.120	0.119	0.119	0.118	0.120	0.129	0.127
25	F.comb.CSR	0.070	0.095	0.100	0.103	0.110	0.126	0.135	0.129	0.141	0.151	0.168	0.183
26	F.comb.Adalasso	0.069	0.091	0.108	0.119	0.116	0.145	0.169	0.170	0.193	0.203	0.222	0.195
27	F.comb.RF	0.075	0.091	0.101	0.110	0.133	0.135	0.157	0.144	0.173	0.161	0.202	0.183
28	BEI	0.036 ***	0.065 **	0.074 **	0.080 **	0.085 *	0.101	0.102	0.107	0.116	0.121	0.128	0.122
29	Focus	0.048 **	0.073 **	0.081 *	0.085 *	0.088 **	0.093 *	0.095 *	0.099 *	0.102	0.104	0.105	0.104
30	F.comb.BEI GR	0.032 ***	0.075 **	0.091	0.112	0.123	0.182	0.142	0.141	0.173	0.177	0.197	0.195
31	F.comb.BEI CLS	0.039 ***	0.071 **	0.077 *	0.085 *	0.093 *	0.108	0.107	0.114	0.116	0.115	0.120	0.120
32	F.comb.BEI CSR	0.056 ***	0.081 **	0.089	0.095	0.102	0.120	0.126	0.123	0.135	0.145	0.161	0.172
33	F.comb.BEI Adalasso	0.035 ***	0.066 **	0.082	0.103	0.108	0.130	0.113	0.122	0.149	0.169	0.185	0.164
34	F.comb.BEI RF	0.055 **	0.080 **	0.093	0.105	0.117	0.123	0.138	0.139	0.162	0.155	0.189	0.172
35	*BEI critical date*	*0.017*	*0.047*	*0.067*	*0.076*	*0.082*	*0.089*	*0.104*	*0.106*	*0.111*	*0.119*	*0.125*	*0.130*
36	*Focus critical date*	*0.016*	*0.053*	*0.076*	*0.082*	*0.086*	*0.091*	*0.093*	*0.097*	*0.099*	*0.103*	*0.105*	*0.104*
	Best model	***30***	***28***	***28***	***28***	***28***	***29***	***29***	***29***	***29***	***29***	***29***	***29***

Empirical Exercise - Results

Figure 2 - Variable selection, Adalasso ($h = 1$)

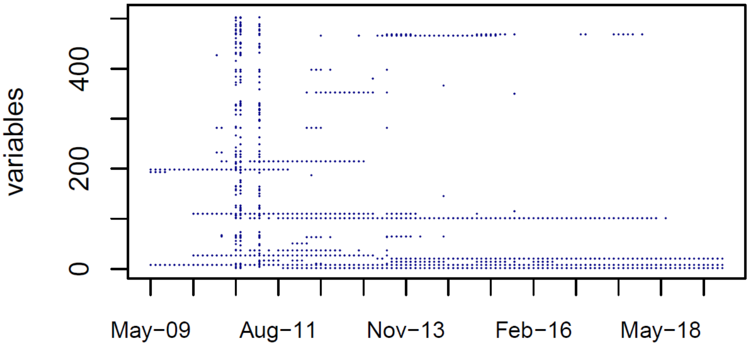
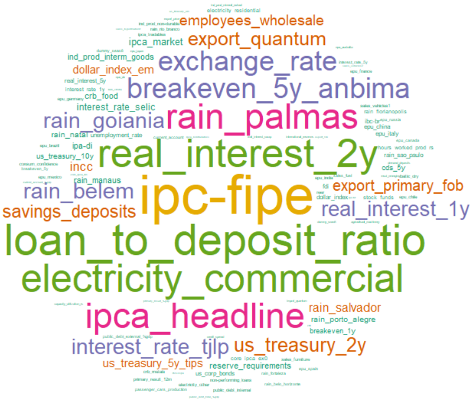


Figure 3 - Word Cloud, Adalasso ($h = 1, 2, 3$)



Conclusion

- Forecasts of IPCA monthly inflation from 36 competing methods.
- Some ML methods yield a sizeable reduction in the forecast variance, while keeping the forecast bias under control.
- As result, forecast accuracy can be improved over traditional models, thus offering a relevant addition to the field of macro forecasting.
- Next steps: Neural networks (LSTM), hybrid models.