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

$$y_{t+h} = \Phi_h (\tilde{x}_t) + \varepsilon_{t+h},$$

(1)

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

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

$$y_{t+h} = \tilde{x}_t' \beta_h + \varepsilon_{t+h}.$$  

(2)
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 $\lambda$, and a combination parameter $\alpha \in [0, 1]$, the elastic net solves:

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

where

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

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


- 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:

$$x_{i,t} = \lambda_i F_t + e_{i,t},$$

where $F_t$ is a vector of common factors and $\lambda_i$ is a vector of loadings. Direct forecast approach:

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

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

$$y_t = \gamma F_t + \varepsilon_t,$$

and $F_t$ follows a VAR process (Bańbura et al., 2013).
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 $E (y_{t+h} - y_t \mid \mathcal{F}_t) = 0$.

**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), $\Delta \ln(M4)$, and $\Delta \ln(FX\text{-rate})$.

**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)
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 $h = 1, \ldots, 12$ 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)

<table>
<thead>
<tr>
<th>Model</th>
<th>Name</th>
<th>h1</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
<th>h5</th>
<th>h6</th>
<th>h7</th>
<th>h8</th>
<th>h9</th>
<th>h10</th>
<th>h11</th>
<th>h12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RW</td>
<td>0.093</td>
<td>0.152</td>
<td>0.160</td>
<td>0.185</td>
<td>0.223</td>
<td>0.211</td>
<td>0.210</td>
<td>0.170</td>
<td>0.161</td>
<td>0.176</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>RW-AO</td>
<td>0.121</td>
<td>0.125</td>
<td>0.127</td>
<td>0.127</td>
<td>0.129</td>
<td>0.132</td>
<td>0.135</td>
<td>0.139</td>
<td>0.143</td>
<td>0.145</td>
<td>0.145</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>AR</td>
<td>0.076</td>
<td>0.106</td>
<td>0.109</td>
<td>0.115</td>
<td>0.122</td>
<td>0.122</td>
<td>0.124</td>
<td>0.125</td>
<td>0.124</td>
<td>0.126</td>
<td>0.129</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>VAR</td>
<td>0.075</td>
<td>0.110</td>
<td>0.111</td>
<td>0.125</td>
<td>0.123</td>
<td>0.122</td>
<td>0.123</td>
<td>0.124</td>
<td>0.124</td>
<td>0.128</td>
<td>0.129</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PC backward</td>
<td>0.075</td>
<td>0.111</td>
<td>0.121</td>
<td>0.125</td>
<td>0.157</td>
<td>0.160</td>
<td>0.160</td>
<td>0.145</td>
<td>0.124</td>
<td>0.121</td>
<td>0.125</td>
<td>0.120</td>
</tr>
<tr>
<td>6</td>
<td>PC hybrid</td>
<td>0.072</td>
<td>*</td>
<td>0.100</td>
<td>0.119</td>
<td>0.132</td>
<td>0.169</td>
<td>0.179</td>
<td>0.186</td>
<td>0.163</td>
<td>0.139</td>
<td>0.130</td>
<td>0.128</td>
</tr>
<tr>
<td>7</td>
<td>Factor model1</td>
<td>0.069</td>
<td>0.096</td>
<td>*</td>
<td>0.105</td>
<td>0.108</td>
<td>0.123</td>
<td>0.134</td>
<td>0.126</td>
<td>0.123</td>
<td>0.124</td>
<td>0.119</td>
<td>0.134</td>
</tr>
<tr>
<td>8</td>
<td>Factor model2</td>
<td>0.067</td>
<td>0.087</td>
<td>*</td>
<td>0.102</td>
<td>0.112</td>
<td>0.116</td>
<td>0.116</td>
<td>0.113</td>
<td>0.115</td>
<td>0.116</td>
<td>0.121</td>
<td>0.125</td>
</tr>
<tr>
<td>9</td>
<td>Factor model3</td>
<td>0.063</td>
<td>0.080</td>
<td>*</td>
<td>0.093</td>
<td>0.102</td>
<td>0.103</td>
<td>0.126</td>
<td>0.143</td>
<td>0.126</td>
<td>0.122</td>
<td>0.123</td>
<td>0.132</td>
</tr>
<tr>
<td>10</td>
<td>Factor model4</td>
<td>0.073</td>
<td>0.093</td>
<td>0.091</td>
<td>*</td>
<td>0.100</td>
<td>0.110</td>
<td>0.111</td>
<td>0.113</td>
<td>0.114</td>
<td>0.117</td>
<td>0.123</td>
<td>0.126</td>
</tr>
<tr>
<td>11</td>
<td>Elastic net</td>
<td>0.079</td>
<td>0.106</td>
<td>0.100</td>
<td>0.102</td>
<td>0.096</td>
<td>**</td>
<td>0.101</td>
<td>**</td>
<td>0.122</td>
<td>0.129</td>
<td>0.146</td>
<td>0.130</td>
</tr>
<tr>
<td>12</td>
<td>Lasso</td>
<td>0.078</td>
<td>0.101</td>
<td>0.099</td>
<td>0.097</td>
<td>0.100</td>
<td>**</td>
<td>0.101</td>
<td>*</td>
<td>0.127</td>
<td>0.136</td>
<td>0.147</td>
<td>0.139</td>
</tr>
<tr>
<td>13</td>
<td>Adalasso</td>
<td>0.077</td>
<td>0.098</td>
<td>0.099</td>
<td>0.103</td>
<td>0.103</td>
<td>*</td>
<td>0.102</td>
<td>**</td>
<td>0.141</td>
<td>0.135</td>
<td>0.126</td>
<td>0.143</td>
</tr>
<tr>
<td>14</td>
<td>Ridge regression</td>
<td>0.073</td>
<td>0.092</td>
<td>*</td>
<td>0.093</td>
<td>0.094</td>
<td>*</td>
<td>0.102</td>
<td>*</td>
<td>0.110</td>
<td>0.113</td>
<td>0.121</td>
<td>0.120</td>
</tr>
<tr>
<td>15</td>
<td>Random forest</td>
<td>0.073</td>
<td>0.089</td>
<td>**</td>
<td>0.092</td>
<td>**</td>
<td>0.097</td>
<td>**</td>
<td>0.104</td>
<td>**</td>
<td>0.111</td>
<td>0.115</td>
<td>0.117</td>
</tr>
<tr>
<td>16</td>
<td>Quant.reg.forest</td>
<td>0.072</td>
<td>0.089</td>
<td>**</td>
<td>0.092</td>
<td>**</td>
<td>0.097</td>
<td>**</td>
<td>0.103</td>
<td>**</td>
<td>0.109</td>
<td>0.115</td>
<td>0.117</td>
</tr>
<tr>
<td>17</td>
<td>Disag. ARFIMA</td>
<td>0.081</td>
<td>0.108</td>
<td>0.114</td>
<td>0.117</td>
<td>0.123</td>
<td>0.122</td>
<td>0.125</td>
<td>0.126</td>
<td>0.124</td>
<td>0.127</td>
<td>0.131</td>
<td>0.133</td>
</tr>
<tr>
<td>18</td>
<td>Disag. Adalasso</td>
<td>0.091</td>
<td>0.096</td>
<td>0.101</td>
<td>0.102</td>
<td>0.108</td>
<td>0.107</td>
<td>0.109</td>
<td>0.108</td>
<td>0.118</td>
<td>0.122</td>
<td>0.123</td>
<td>0.129</td>
</tr>
<tr>
<td>19</td>
<td>Disag. RF</td>
<td>0.072</td>
<td>*</td>
<td>0.091</td>
<td>*</td>
<td>0.094</td>
<td>*</td>
<td>0.097</td>
<td>*</td>
<td>0.104</td>
<td>*</td>
<td>0.110</td>
<td>0.117</td>
</tr>
<tr>
<td>20</td>
<td>F.comb.Median</td>
<td>0.070</td>
<td>0.093</td>
<td>0.095</td>
<td>*</td>
<td>0.089</td>
<td>**</td>
<td>0.106</td>
<td>**</td>
<td>0.109</td>
<td>**</td>
<td>0.116</td>
<td>0.120</td>
</tr>
<tr>
<td>21</td>
<td>F.comb.Mean</td>
<td>0.068</td>
<td>0.088</td>
<td>**</td>
<td>0.093</td>
<td>0.097</td>
<td>**</td>
<td>0.107</td>
<td>**</td>
<td>0.109</td>
<td>**</td>
<td>0.116</td>
<td>0.116</td>
</tr>
<tr>
<td>22</td>
<td>F.comb.Mean2</td>
<td>0.068</td>
<td>0.088</td>
<td>**</td>
<td>0.092</td>
<td>**</td>
<td>0.093</td>
<td>0.102</td>
<td>**</td>
<td>0.106</td>
<td>**</td>
<td>0.118</td>
<td>0.118</td>
</tr>
<tr>
<td>23</td>
<td>F.comb.GR</td>
<td>0.067</td>
<td>*</td>
<td>0.097</td>
<td>0.113</td>
<td>0.127</td>
<td>0.136</td>
<td>0.198</td>
<td>0.170</td>
<td>0.164</td>
<td>0.203</td>
<td>0.209</td>
<td>0.216</td>
</tr>
<tr>
<td>24</td>
<td>F.comb.CLS</td>
<td>0.063</td>
<td>0.086</td>
<td>**</td>
<td>0.091</td>
<td>0.083</td>
<td>0.099</td>
<td>0.120</td>
<td>0.119</td>
<td>0.119</td>
<td>0.118</td>
<td>0.120</td>
<td>0.129</td>
</tr>
<tr>
<td>25</td>
<td>F.comb.CSR</td>
<td>0.070</td>
<td>0.095</td>
<td>0.100</td>
<td>0.103</td>
<td>0.110</td>
<td>0.126</td>
<td>0.135</td>
<td>0.129</td>
<td>0.141</td>
<td>0.151</td>
<td>0.168</td>
<td>0.183</td>
</tr>
<tr>
<td>26</td>
<td>F.comb.Adalasso</td>
<td>0.069</td>
<td>0.091</td>
<td>0.108</td>
<td>0.119</td>
<td>0.116</td>
<td>0.145</td>
<td>0.169</td>
<td>0.170</td>
<td>0.193</td>
<td>0.203</td>
<td>0.222</td>
<td>0.195</td>
</tr>
<tr>
<td>27</td>
<td>F.comb.RF</td>
<td>0.075</td>
<td>0.091</td>
<td>0.101</td>
<td>0.110</td>
<td>0.133</td>
<td>0.135</td>
<td>0.157</td>
<td>0.144</td>
<td>0.173</td>
<td>0.161</td>
<td>0.202</td>
<td>0.183</td>
</tr>
</tbody>
</table>

**Best model** 24 9 10 22 11 11 18 18 21 21 16 6
### Table 2 - Mean Squared Error (MSE)

<table>
<thead>
<tr>
<th>Model</th>
<th>Name</th>
<th>h1</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
<th>h5</th>
<th>h6</th>
<th>h7</th>
<th>h8</th>
<th>h9</th>
<th>h10</th>
<th>h11</th>
<th>h12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RW</td>
<td>0.093</td>
<td>0.152</td>
<td>0.160</td>
<td>0.185</td>
<td>0.223</td>
<td>0.211</td>
<td>0.219</td>
<td>0.210</td>
<td>0.170</td>
<td>0.161</td>
<td>0.176</td>
<td>0.191</td>
</tr>
<tr>
<td>2</td>
<td>RW-AO</td>
<td>0.121</td>
<td>0.125</td>
<td>0.127</td>
<td>0.127</td>
<td>0.129</td>
<td>0.132</td>
<td>0.135</td>
<td>0.139</td>
<td>0.139</td>
<td>0.143</td>
<td>0.145</td>
<td>0.145</td>
</tr>
<tr>
<td>3</td>
<td>AR</td>
<td>0.076</td>
<td>0.106</td>
<td>0.109</td>
<td>0.115</td>
<td>0.122</td>
<td>0.122</td>
<td>0.124</td>
<td>0.124</td>
<td>0.125</td>
<td>0.126</td>
<td>0.129</td>
<td>0.128</td>
</tr>
<tr>
<td>4</td>
<td>VAR</td>
<td>0.075</td>
<td>0.110</td>
<td>0.111</td>
<td>0.125</td>
<td>0.125</td>
<td>0.122</td>
<td>0.122</td>
<td>0.122</td>
<td>0.124</td>
<td>0.124</td>
<td>0.128</td>
<td>0.129</td>
</tr>
<tr>
<td>5</td>
<td>PC backward</td>
<td>0.075</td>
<td>0.111</td>
<td>0.121</td>
<td>0.125</td>
<td>0.157</td>
<td>0.160</td>
<td>0.160</td>
<td>0.145</td>
<td>0.124</td>
<td>0.121</td>
<td>0.125</td>
<td>0.120</td>
</tr>
<tr>
<td>6</td>
<td>PC hybrid</td>
<td>0.072</td>
<td>0.100</td>
<td>0.119</td>
<td>0.132</td>
<td>0.169</td>
<td>0.179</td>
<td>0.186</td>
<td>0.163</td>
<td>0.139</td>
<td>0.130</td>
<td>0.128</td>
<td>0.117</td>
</tr>
</tbody>
</table>

#### Empirical Exercise - Results

**Best model**

<table>
<thead>
<tr>
<th>Model</th>
<th>Name</th>
<th>h1</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
<th>h5</th>
<th>h6</th>
<th>h7</th>
<th>h8</th>
<th>h9</th>
<th>h10</th>
<th>h11</th>
<th>h12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RW</td>
<td>0.093</td>
<td>0.152</td>
<td>0.160</td>
<td>0.185</td>
<td>0.223</td>
<td>0.211</td>
<td>0.219</td>
<td>0.210</td>
<td>0.170</td>
<td>0.161</td>
<td>0.176</td>
<td>0.191</td>
</tr>
<tr>
<td>2</td>
<td>RW-AO</td>
<td>0.121</td>
<td>0.125</td>
<td>0.127</td>
<td>0.127</td>
<td>0.129</td>
<td>0.132</td>
<td>0.135</td>
<td>0.139</td>
<td>0.139</td>
<td>0.143</td>
<td>0.145</td>
<td>0.145</td>
</tr>
<tr>
<td>3</td>
<td>AR</td>
<td>0.076</td>
<td>0.106</td>
<td>0.109</td>
<td>0.115</td>
<td>0.122</td>
<td>0.122</td>
<td>0.124</td>
<td>0.124</td>
<td>0.125</td>
<td>0.126</td>
<td>0.129</td>
<td>0.128</td>
</tr>
<tr>
<td>4</td>
<td>VAR</td>
<td>0.075</td>
<td>0.110</td>
<td>0.111</td>
<td>0.125</td>
<td>0.125</td>
<td>0.122</td>
<td>0.122</td>
<td>0.122</td>
<td>0.124</td>
<td>0.124</td>
<td>0.128</td>
<td>0.129</td>
</tr>
<tr>
<td>5</td>
<td>PC backward</td>
<td>0.075</td>
<td>0.111</td>
<td>0.121</td>
<td>0.125</td>
<td>0.157</td>
<td>0.160</td>
<td>0.160</td>
<td>0.145</td>
<td>0.124</td>
<td>0.121</td>
<td>0.125</td>
<td>0.120</td>
</tr>
<tr>
<td>6</td>
<td>PC hybrid</td>
<td>0.072</td>
<td>0.100</td>
<td>0.119</td>
<td>0.132</td>
<td>0.169</td>
<td>0.179</td>
<td>0.186</td>
<td>0.163</td>
<td>0.139</td>
<td>0.130</td>
<td>0.128</td>
<td>0.117</td>
</tr>
</tbody>
</table>

**17 / 20**
Empirical Exercise - Results

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

Araujo and Gaglianone (2020) ()
ML methods for inflation forec. in Brazil
Empirical Exercise - Results

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.