Inflation Forecasting Utilizing High Dimensionality Models

Marco Aurélio Guerra

Data Gathering

For the Forecast of the monthly Brazilian CPI (IPCA), I used 70 variables[1], With these data used from 3 sources, Brazilian Central Bank Database (R packages GetBCBdata, rbcb), Yahoo Finance (BatchGetSymbols), and National Treasury Results Bulletin[2].

Data Treatment

The variation and expectation of the monthly CPI, following the methodology used in Figueiredo (2010) and Medeiros et al. (2019), were transformed into log return[3]. For the data collected that had a daily frequency, the last data observed in the month was considered[4].

In the case of expectations, it was necessary to unstack the data for the different forecast horizons.

To avoid the problem of non-stationarity, an augmented dick-fuller test was performed considering the trend and drift of the series. There being no rejection of the null hypothesis of non-stationarity[5] the data were transformed into log difference[6] .

Modeling

The modeling used in this work is based on two articles. The article written by Garcia, Medeiros, and Vasconcelos (2017) applies high-dimensional models to forecast inflation in Brazil. The work of Medeiros et al. (2019) expands the number of models used in Garcia, Medeiros, and Vasconcelos (2017)'s text and applies them to estimate the price index of the North American economy. Both articles provided the codes used in the form of a package for the R[7] [8].

The monthly IPCA forecasts are *out of sample* estimates for 108 months, where the model is re-estimated at each period (*rolling window*) with a window of 106 months, observing the period from 2003-01 to 2020-12. In all models, the IPCA estimates were performed for 13 horizons, $h = \{1, 2, ..., 13\}$. Model estimates are evaluated by the root mean square of errors (RMSE)[9] and by the Mean Absolute Error (MAE)[10].

To estimate the monthly variation of the CPI, 14 models and the average of all projections were used. The list of models below was based on the article by Medeiros et al. (2019):

[1] See description in notes

[2] https://www.tesourotransparente.gov.br/publicacoes/boletim-resultado-do-tesouro-nacional-rtn/2021/2.I used the spreadsheet as the API data is not up to date since August 2020.

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

[4] 'Euro sale exchange rate,''Daily balance of savings deposits,' 'Daily net raising of savings deposits,''Average floating rate DI of time deposits (CDB/RDB),''Interest rate - Selic target defined by the Copom' ,'bovespa index'

[5] Tested models had the specification of 1 to 4 lags of the variable

[6] $x_t = \log(X_t) - \log(X_{t-1})$

[7] See https://github.com/gabrielrvsc/ForecastingInflation and https://github.com/gabrielrvsc/HDeconometrics

[8] I used their function scripts individually. Also, some code changes were necessary, as I found some bugs.

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

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

- Benchmark models: AR, Random Walk (RW), Unobserved Stochastic Volatility Component (UCSV)
- Shrinkage models: Lasso, adaLasso, flex-adaLasso
- Factor Models: Factor Model, Target Factors, Factor Boosting
- Ensemble Models: Complete Subset Regressions (CSR)
- Random Forest Models: Random Forest (RF), Random Forest with Ordinary Least Squares, Random Forest with AdaLasso
- Autoregressive Vector Model: Var Bayesian
- Average of predictions from all previous models

Results

The results of all models were stored in a list-type object ('all_forecasts_by_horizon'). With these estimates, the RMSE and MAE were calculated for all models and horizons. To assist in the evaluation of the RMSE and MAE results, the random walk model was used as a reference through the following calculation:

$RelativeScore = \frac{RMSE(\hat{Y}_{model})}{RMSE(\hat{Y}_{RandomWalk})}$

The following table contains the calculated Relative Score results for the RMSE and MAE. The result of this analysis is that the most accurate model, in general, is the average of all models. However, for the first period, the Lasso and Random Forest models with Ordinary Least Squares have a better result than the RMSE.

model	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	0.84	0.80	0.76	0.77	0.76	0.82	0.83	0.93	0.96	0.92	0.91	0.85
	(0.92)	(0.84)	(0.80)	(0.76)	(0.77)	(0.76)	(0.82)	(0.83)	(0.93)	(0.96)	(0.92)	(0.91)	(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	0.96	0.99	0.96	0.96	0.93	0.95	0.93	0.93	0.98	0.96	0.95	0.96
	(1.00)	(0.96)	(0.99)	(0.96)	(0.96)	(0.93)	(0.95)	(0.93)	(0.93)	(0.98)	(0.96)	(0.95)	(0.96)
models													
lasso	0.69	0.76	0.74	0.70	0.72	0.74	0.80	0.78	0.90	0.86	0.85	0.97	0.88
	(0.69)	(0.76)	(0.74)	(0.70)	(0.72)	(0.74)	(0.80)	(0.78)	(0.90)	(0.86)	(0.85)	(0.97)	(0.88)
ridge	1.11	0.87	0.83	0.75	0.75	0.73	0.79	0.80	0.90	0.93	0.89	0.89	0.83
	(1.11)	(0.87)	(0.83)	(0.75)	(0.75)	(0.73)	(0.79)	(0.80)	(0.90)	(0.93)	(0.89)	(0.89)	(0.83)
adalasso	0.69	0.76	0.74	0.70	0.72	0.74	0.80	0.78	0.90	0.86	0.85	0.97	0.88
	(0.69)	(0.76)	(0.74)	(0.70)	(0.72)	(0.74)	(0.80)	(0.78)	(0.90)	(0.86)	(0.85)	(0.97)	(0.88)
flex_adalasso	0.69	0.76	0.74	0.70	0.72	0.74	0.80	0.78	0.90	0.86	0.85	0.97	0.88
	(0.69)	(0.76)	(0.74)	(0.70)	(0.72)	(0.74)	(0.80)	(0.78)	(0.90)	(0.86)	(0.85)	(0.97)	(0.88)
fact	0.96	0.84	0.82	0.79	0.78	0.73	0.75	0.78	0.83	0.92	0.83	0.82	0.79
	(0.96)	(0.84)	(0.82)	(0.79)	(0.78)	(0.73)	(0.75)	(0.78)	(0.83)	(0.92)	(0.83)	(0.82)	(0.79)
target_fact	0.88	0.79	0.79	0.76	0.75	0.71	0.81	0.80	0.91	0.91	0.86	0.84	0.79
	(0.88)	(0.79)	(0.79)	(0.76)	(0.75)	(0.71)	(0.81)	(0.80)	(0.91)	(0.91)	(0.86)	(0.84)	(0.79)
boosting_fact	1.02	0.83	0.81	0.76	0.70	0.68	0.74	0.70	0.77	0.79	0.77	0.78	0.77
	(1.02)	(0.83)	(0.81)	(0.76)	(0.70)	(0.68)	(0.74)	(0.70)	(0.77)	(0.79)	(0.77)	(0.78)	(0.77)
csr	0.75	0.74	0.73	0.68	0.67	0.68	0.74	0.72	0.80	0.85	0.84	0.82	0.75
	(0.75)	(0.74)	(0.73)	(0.68)	(0.67)	(0.68)	(0.74)	(0.72)	(0.80)	(0.85)	(0.84)	(0.82)	(0.75)
rf	0.78	0.75	0.72	0.65	0.66	0.64	0.71	0.71	0.78	0.80	0.78	0.79	0.76

model	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
	(0.78)	(0.75)	(0.72)	(0.65)	(0.66)	(0.64)	(0.71)	(0.71)	(0.78)	(0.80)	(0.78)	(0.79)	(0.76)
rfols	**0.70**	0.74	0.72	0.65	0.64	0.64	0.70	0.71	0.81	0.83	0.81	0.83	0.80
	(0.70)	(0.74)	(0.72)	(0.65)	(0.64)	(0.64)	(0.70)	(0.71)	(0.81)	(0.83)	(0.81)	(0.83)	(0.80)
rf_adalasso	0.80	0.81	0.74	0.69	0.71	0.70	0.79	0.81	0.82	0.82	0.80	0.83	0.80
	(0.80)	(0.81)	(0.74)	(0.69)	(0.71)	(0.70)	(0.79)	(0.81)	(0.82)	(0.82)	(0.80)	(0.83)	(0.80)
lbvar	0.82	0.74	0.74	0.69	0.71	0.71	0.78	0.80	0.90	0.94	0.91	0.91	0.85
	(0.82)	(0.74)	(0.74)	(0.69)	(0.71)	(0.71)	(0.78)	(0.80)	(0.90)	(0.94)	(0.91)	(0.91)	(0.85)
ucsv	1.00	0.96	0.98	0.96	0.96	0.93	0.95	0.93	0.93	0.98	0.96	0.95	0.96
	(1.00)	(0.96)	(0.98)	(0.96)	(0.96)	(0.93)	(0.95)	(0.93)	(0.93)	(0.98)	(0.96)	(0.95)	(0.96)
E(models)	0.72	**0.70**	**0.68**	**0.64**	**0.65**	**0.61**	**0.68**	**0.68**	**0.74**	**0.76**	**0.73**	**0.76**	**0.72**
	(0.72)	(0.70)	(0.68)	(0.64)	(0.65)	(0.61)	(0.68)	(0.68)	(0.74)	(0.76)	(0.73)	(0.76)	(0.72)
FOCUS	0.99	0.77	0.75	0.67	0.68	0.67	0.73	0.75	0.84	0.87	0.83	0.83	0.77
	(0.99)	(0.77)	(0.75)	(0.67)	(0.68)	(0.67)	(0.73)	(0.75)	(0.84)	(0.87)	(0.83)	(0.83)	(0.77)
TOP5FOCUS	1.01	0.78	0.78	0.68	0.68	0.68	0.75	0.77	0.86	0.88	0.83	0.83	0.77
	(1.01)	(0.78)	(0.78)	(0.68)	(0.68)	(0.68)	(0.75)	(0.77)	(0.86)	(0.88)	(0.83)	(0.83)	(0.77)

Note: The numbers in parentheses are the values of the Mean Absolute Error (MAE) the rest are the root mean square of the errors (RMSE).

Bibliography

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.

Notes

Code/Source	Variable Name	Unit	Type
433 - BACEN	IPCA (CPI)	Monthly % variation	Price and Money
189 - BACEN	IGP-M	Monthly % variation	Price and Money
190 - BACEN	IGP-DI	Monthly % variation	Price and Money
7447 - BACEN	IGP-10	Monthly % variation	Price and Money
1788 - BACEN	Restricted monetary base (end-of-period balance).	u.m.c (thousand)	Price and Money
27791 - BACEN	M1 (end-of-period balance) - New	u.m.c (thousand)	Price and Money
27810 - BACEN	M2 (end-of-period balance) - New	u.m.c (thousand)	Price and Money
27813 - BACEN	M3 (end-of-period balance) - New	u.m.c (thousand)	Price and Money

Code/Source	Variable Name	Unit	Type
27815 - BACEN	M4 (end-of-period balance) - New	u.m.c (thousand)	Price and Money
24369 - BACEN	Vacancy rate - PNADC	percentage	employment
24348 - BACEN	Number of hours worked - manufacturing industry (2006=100)	index	employment
28766 - BACEN	Stock of formal jobs - Manufacturing industries	units	employment
3695 - BACEN	Exchange rate - Free - US dollar (buy) - End of period - monthly	R$ u.m.c	FX and Finance
21619 - BACEN	Exchange rate - Free - Euro (for sale)	R$ u.m.c	FX and Finance
Yahoo Finance	Ibov	Points	FX and Finance
23 - BACEN	Daily balance of savings deposits - SBPE and rural	u.m.c (thousand)	FX and Finance
24 - BACEN	Daily net raising of savings deposits - SBPE and rural	u.m.c (thousand)	FX and Finance
256 - BACEN	Long-term interest rate - TJLP	per. a.a.	FX and Finance
432 - BACEN	Interest rate - Selic target defined by Copom	per. a.a.	FX and Finance
1157 - BACEN	Average floating rate DI on time deposits (CDB/RDB) - Total	per. a.d.	FX and Finance
22701 - BACEN	Current Account - monthly - balance	US$ (millions)	Government and international transactions
22707 - BACEN	Balance of trade - Balance of Payments - monthly - balance	US$ (millions)	Government and international transactions
4503 - BACEN	Net Public Debt as a Percentage of GDP	Porcentagem	Government and international transactions
10825 - BACEN	Net Public Debt	US$ (millions)	Government and international transactions
4480 - BACEN	Internal debt of the Central Government	US$ (millions)	Government and international transactions
4479 - BACEN	Net Internal Debt of the Central Government and Central Bank	US$ (millions)	Government and international transactions
4472 - BACEN	Net Debt of States	US$ (millions)	Government and international transactions
4494 - BACEN	Net Debt of States with foreigners	US$ (millions)	Government and international transactions
4473 - BACEN	Municipalities' Net Debt	US$ (millions)	Government and international transactions

Code/Source	Variable Name	Unit	Type
4495 - BACEN	Municipalities' Net Debt with foreigners	US$ (millions)	Government and international transactions
Treasury spreadsheet	PIS & PASEP Revenue	US$ (millions)	Government and international transactions
Treasury spreadsheet	central government revenue	US$ (millions)	Government and international transactions
Treasury spreadsheet	Total Central Government Spending	US$ (millions)	Government and international transactions
Treasury spreadsheet	Income from Import Tax	US$ (millions)	Government and international transactions
Treasury spreadsheet	Income from Import Tax	US$ (millions)	Government and international transactions
Treasury spreadsheet	Revenue from other taxes	US$ (millions)	Government and international transactions
Treasury spreadsheet	Social Security System Revenue	US$ (millions)	Government and international transactions
BACEN	Median of market expectations for IPCA in t (for t from 1 to 13)	Monthly % variation do Índice	Forecast
BACEN	standard deviation of market expectations (for t equal to 1, 2 and 12)	NA	Forecast
BACEN	Median of TOP5 Market Expectations for IPCA in t (for t from 1 to 13)	Monthly % variation do Índice	Forecast