Macroeconometrics and Machine Learning Project

Louis Blanco & Jeanne Astier

26 décembre 2022

This project is based on the article "Forecasting Inflation in a Data-Rich Environment : The Benefits of Machine Learning Methods" by Marcelo C. Medeiros, Gabriel F. R. Vasconcelos, Álvaro Veiga and Eduardo Zilberman (2019). It is organized as follows : part 1 explains the problem under study and its interests. Part 2 describes our replication of the article results in forecasting US inflation. Part 3 presents an extension of these results, applying the same methodology to forecast French inflation. Part 4 provides economic interpretation of the results and methodology used.

1 Introduction

Price Gains Slow More Than Expected : this is the title of the November Inflation Report published by the New York Times on December 13$^{\text{th}}$ 2022. It highlights the complexity of inflation forecasting : after a year during which the Fed repeatedly underestimated inflation's hold on the U.S. [1], it seems now to be overestimating it. These forecast errors, common to many Central Banks (the ECB addresses in its Economic Bulletin 2022 its recent errors in the inflation projections), are particularly problematic as inflation forecast is of paramount importance for economic decisions. Central Banks rely on inflation forecasts to inform their monetary policy. These forecasts are also used by households, private sector companies, policymakers to build their expectations and make their decisions. In some macroeconomic models, not only realized inflation but also expectations of inflation play a central role. Hence there is a discrepancy between the prime importance of inflation forecasts and the their lack of accuracy. These forecasts errors also undermine Central Banks credibility, degrading again the expectations. Poor inflation forecasts might thus have important welfare costs.

1. *2022 shattered economic forecasts. Can the Fed get 2023 right ?*, The Washington Post, December 12$^{\text{th}}$ 2022.

In this context, there is a challenge to improve performance of forecasting models. According to literature, the traditional simple univariate forecasting models turn out to be difficult to improve. Even the attempts to use Machine Learning models on a small set of variables, or "Big data" summarized in a small set of factors, did not seem conclusive. However, combining the two, as it has already been done outside of the economic field, might lead to significant improvement. The objective of the article studied is not only to beat univariate benchmark models, but also to prove that the combination of Machine Learning and "Big data" is the more effective that using just Machine Learning with few variables of Big Data with factor models.

2 Replication of article results

We replicate the article results by estimating benchmark models and several Machine Learning ones on a large set of variables to predict US inflation. We then compare the performances of the forecasts made by these different models, to determine which ones are the most accurate.

We face several obstacles in this replication, which lead us to make choices in how to run it. The main one is the computing power required to run all the models compared. Not only the machine learning models are power-consuming to be trained, but also all the processes are replicated many times (forecasts are computed at 12 different time horizons at each one of the 312 out-of-sample windows). This factor is highly limiting since our laptops do not have very high computing power, so the predictions can take very long to be computed [2]. This explains our choice to stick to the replication of general results of the paper, and not to replicate exhaustively all the robustness checks made by the authors that would have asked several days of code running with our tools. For this reason, we also gave up on replicating forecasts for 4 models among the 18 tested.

An other issue is that the data available to us do not seem to correspond exactly to what the authors describe. Even sticking to their description of the source of the data (January 2016 edition of FRED-MD data) and to the treatment they apply (among others, removing the variables with missing dates in the time period considered), we end up with slightly less variables than they do, without being able to identify which ones and the reasons for this. This might explain why our results statistics do not always exactly correspond to the article original ones.

As in the original article, we train several models on FRED-MD data from 1960 to 1990 ; we compare performance of these different models in an out-of-sample window from January 1990 to

2. The total computing time for the replication results presented is about 35 hours.

December 2015. At each date t of the out-of-sample window, the compute forecasts based on a rolling-window framework of fixed length, at time horizons $t+1$, $t+1$, ..., $t+12$. We also compute forecasts for the accumulated inflation over the following 3, 6, and 12 months. The models are compared according to different statistics : root mean squared error (RMSE) and mean absolute error (MAE). We also compute, for square and absolute losses, respectively, the average p-values (accross horizons) for the model confidence sets (MCSs) based on the Tmax statistic as described in Hansen, Lunde, and Nason (2011). The estimated models are the following ones :

- benchmark models : Random Walk (RW), Auto-Regressive (AR)
- shrinkage models : Ridge regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO), adaptive LASSO (adaLASSO), Elastic Net (ElNet)
- factor models : classic factor model (Factor), Target factor model (T. Factor)
- ensemble methods : Bagging, Complete Subset Regressions (CSR)
- Breiman Random Forest (RF)
- hybrid linear-Random Forest models : Random-Forest Ordinary Least Squares (RF/OLS).

3 Extension : application to forecasting of French inflation

We try to apply the benchmark and Machine Learning models implemented in the article to French data. There is a double issue at stake with this extension. First, is the domination of Random Forest over all other models specific to US data, or does it hold on a different dataset ? Second, are the Machine Learning models still performing better than benchmarks with a smaller number of covariates ? Indeed, we cannot not find as many monthly series available on a large time period for France as we have for the US with FRED-MD dataset. We download monthly series about France from FRED database, but way less series are available than for the US, and additionally most of them start later than 1960. We will thus see how much "Big Data" is needed for Machine Learning models to outperform univariate ones.

4 Interpretation of the results

Economic interpretation of the results and explanation of the methodology used

Table 1 – Replication of forecasting results : summary statistics for the out-of-sample period from 1990 to 2015

	Forecasting precision								Model confidence set	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Models	ave. RMSE	ave. MAE	max RMSE	max MAE	min RMSE	min MAE	# min RMSE	# min MAE	ave. p-v. Tmax sq	ave. p-v. Tmax abs
RW	1.00	1.00	1.00	1.00	1.00	1.00	0	0	0.00	0.00
AR	0.61	0.81	0.88	0.87	0.56	0.70	2	0	1.00	0.32
LASSO	0.75	0.94	0.84	1.01	0.62	0.67	3	0	0.28	0.32
adaLASSO	0.85	0.96	0.88	1.04	0.64	0.66	0	0	0.07	0.13
ElNet	0.76	0.87	0.81	1.05	0.62	0.68	1	1	0.26	0.32
adaElNet	0.85	0.98	0.88	1.06	0.65	0.69	1	1	0.08	0.13
Ridge	0.79	0.89	0.86	0.92	0.73	0.71	0	1	0.05	0.13
BVAR	0.62	0.82	0.90	0.89	0.57	0.70	0	0	0.32	0.32
Bagging	1.38	1.25	1.50	1.35	0.82	0.83	0	0	0.05	0.05
CSR	0.74	0.90	0.82	0.95	0.65	0.69	0	0	0.07	0.13
Factor	0.73	0.89	0.83	1.01	0.58	0.66	1	2	0.29	0.32
T. Factor	0.65	0.80	0.84	0.85	0.58	0.66	1	0	0.32	0.32
RF	0.69	0.80	0.83	0.83	0.66	0.71	5	9	0.29	1.00
RF/OLS	0.77	0.87	0.81	0.91	0.64	0.71	1	1	0.20	0.32

NOTE : The table reports for each model different summary statistics across all the forecasting horizons (1 to 12 months and accumulated 3, 6 and 12-month). Columns (1) and (2) report the average root mean square error (RMSE) and the average mean absolute error (MAE). Columns (3) ans (4) report, respectively, the maximum RMSE and MAE over the horizons considered. Columns (5) and (6) report, respectively, the minimum RMSE and MAE over the horizons considered. Ass these statistics (columns 1 to 6) are normalized for the benchmark RW model to one. Columns (7) and (8) report the number of times (across horizons) each model achieved the lowest RMSE and MAE, respectively. Columns (9) and (10) present for square and absolute losses, the average p-values for the model confidence sets (MCSs) based on the Tmax statistic as described in Hansen, Lunde, and Nason (2011).

Table 2 – Replication of forecasting results : RMSE and MAE ratios (1990–2015)

Panel (a) : RMSE ratio

	Forecasting horizon														
Model	1	2	3	4	5	6	7	8	9	10	11	12	3m	6m	12m
AR	0,88	0,80	0,78	0,79	0,77	0,77	0,76	0,74	0,75	0,81	0,81	0,74	**0,64**	**0,66**	**0,56**
RF	**0,84**	**0,75**	**0,73**	**0,76**	**0,73**	**0,74**	**0,75**	**0,74**	**0,74**	**0,78**	**0,80**	**0,72**	0,66	0,73	0,67

Panel (b) : MAE ratio

	Forecasting horizon														
Model	1	2	3	4	5	6	7	8	9	10	11	12	3m	6m	12m
AR	0,87	0,78	0,77	0,78	0,79	0,78	0,75	0,74	0,78	0,83	0,85	0,75	**0,70**	**0,80**	0,82
RF	**0,81**	**0,72**	**0,71**	**0,74**	**0,74**	**0,75**	**0,73**	**0,72**	**0,75**	**0,77**	**0,81**	**0,73**	0,71	0,84	**0,81**

NOTE : The table reports, for each forecasting horizon (1 to 12 months and accumulated 3, 6 and 12 months), the root mean squared error (RMSE) and mean absolute error (MAE) ratios with respect to the random walk model for the full out-of-sample period (1990–2015). The statistics for the best-performing model are highlighted in bold.