Online Appendix to

Harnessing Machine Learning for Real-Time Inflation Nowcasting

Richard Schnorrenberger
Kiel University

Aishameriane Schmidt
Erasmus Universiteit Rotterdam,
Tinbergen Institute and De Nederlandsche Bank

Guilherme Valle Moura
Federal University of Santa Catarina

This version: January 2025

Contents

A Mixed-frequency framework in matrix form

For expositional simplicity, let us reduce the general multiple-predictors specification (1) to the single generic high-frequency predictor $x_t^{(w)}$ and neglect both the low-frequency predictors and seasonal dummies. From there, assume the latest data release for the target variable is associated with a given month t. Based on the high-frequency information set available up to the nowcast point, say $t+1-h$, and pre-sample information $\{\pi_0, x_0^{(w)}, x_{0-1/4}^{(w)}, \ldots, x_{0-p/4}^{(w)}\}$, one can construct the nowcast for π_{t+1} at horizon $h = j/w$, with $j \in \{0,1,2,3\}$, by using the following matrix representation for model estimation:

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_t \end{bmatrix} = \begin{bmatrix} 1 & \pi_0 & x_{1-h}^{(w)} & x_{1-h-\frac{1}{4}}^{(w)} & x_{1-h-\frac{2}{4}}^{(w)} & \cdots & x_{1-h-\frac{p}{4}}^{(w)} \\ 1 & \pi_1 & x_{2-h}^{(w)} & x_{2-h-\frac{1}{4}}^{(w)} & x_{2-h-\frac{2}{4}}^{(w)} & \cdots & x_{2-h-\frac{p}{4}}^{(w)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \pi_{t-1} & \underbrace{x_{t-h}^{(w)}}_{\text{nowcast day } (nd)} & \underbrace{x_{t-h-\frac{1}{4}}^{(w)}}_{nd-\frac{1}{4}} & \underbrace{x_{t-h-\frac{2}{4}}^{(w)}}_{nd-\frac{2}{4}} & \cdots & \underbrace{x_{t-h-\frac{p}{4}}^{(w)}}_{nd-\frac{p}{4}} \end{bmatrix} \begin{bmatrix} c \\ \rho_1 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_{p+1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix} \tag{A1}$$

For example, suppose we stand at day 15 of December and we want to construct the nowcast for π_{Dec} assuming a general high-frequency lag order p. In this case, the forecast horizon is $h = 2/4$ and we estimate the model using monthly data until November and weekly data until 15 November. To account for the lags p, the last high-frequency observations in (A1) will respectively be associated with 15 November, 8 November, 31 October, 22 October, 15 October, and so on up to the corresponding lag-length p. From there, the nowcast for π_{Dec} is constructed using the estimated coefficients and all the low- and high-frequency information available until 15 December.

Ultimately, note that (A1) makes explicit that the general prediction model is still written at the monthly frequency but accounting for the w high-frequency time increments within each common period t. The nowcast for the inflation rate at periods $t+1, \ldots, T$ can then be updated regularly using the high-frequency data increments that become available after t and well before official releases of the target inflation rate.

B Supplementary results

Table B1: RMSE: absolute values

Horizon	SPF		Shrinkage-based methods				Tree-based methods				Naive Models	
	Median	Top 5	LASSO	Ridge	Enet	sg-LASSO	RF	LLF	BART	LASSO-LLF	AR	RW
day 8	0,251	0,235**	0,215***	0,222**	0,218**	0,239	0,257	0,285	0,247	0,241	0,341	0,460
day 15	0,196	0,199	0,184	0,189	0,185	0,201	0,233	0,214	0,218	0,206	0,341	0,378
day 22	0,143	0,136**	0,128**	0,135	0,128**	0,164	0,191	0,155	0,173	0,132	0,341	0,378
End-of-month	0,119	0,115	0,110*	0,117	0,110*	0,113	0,167	0,142	0,154	0,145	0,341	0,378

Notes: The table reports the RMSE for each competing model. Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the benchmark are indicated by the symbols * (1% level), ** (5% level) and *** (10% level).

Figure B1: CUMSFE: tree-based methods versus the SPF benchmark

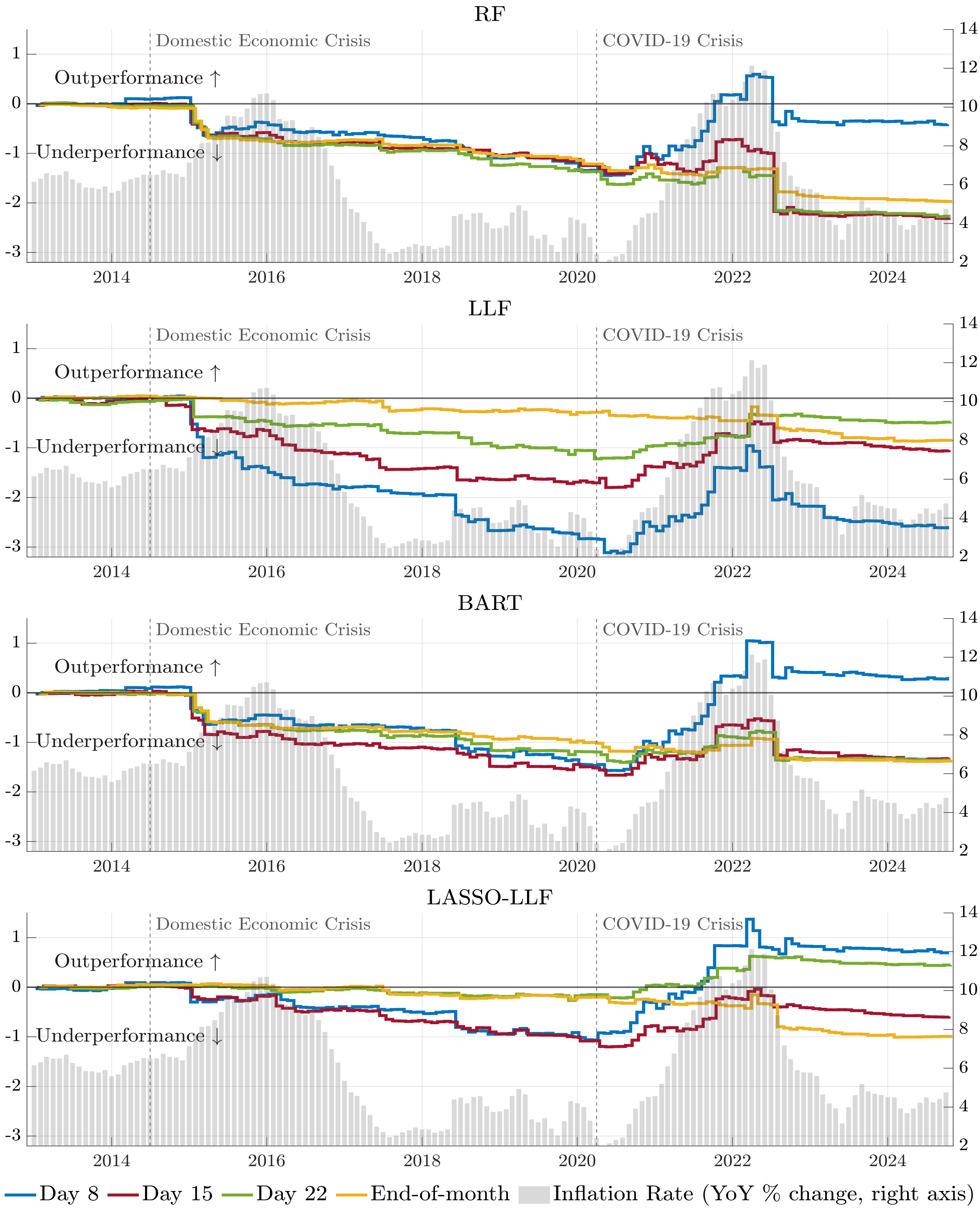
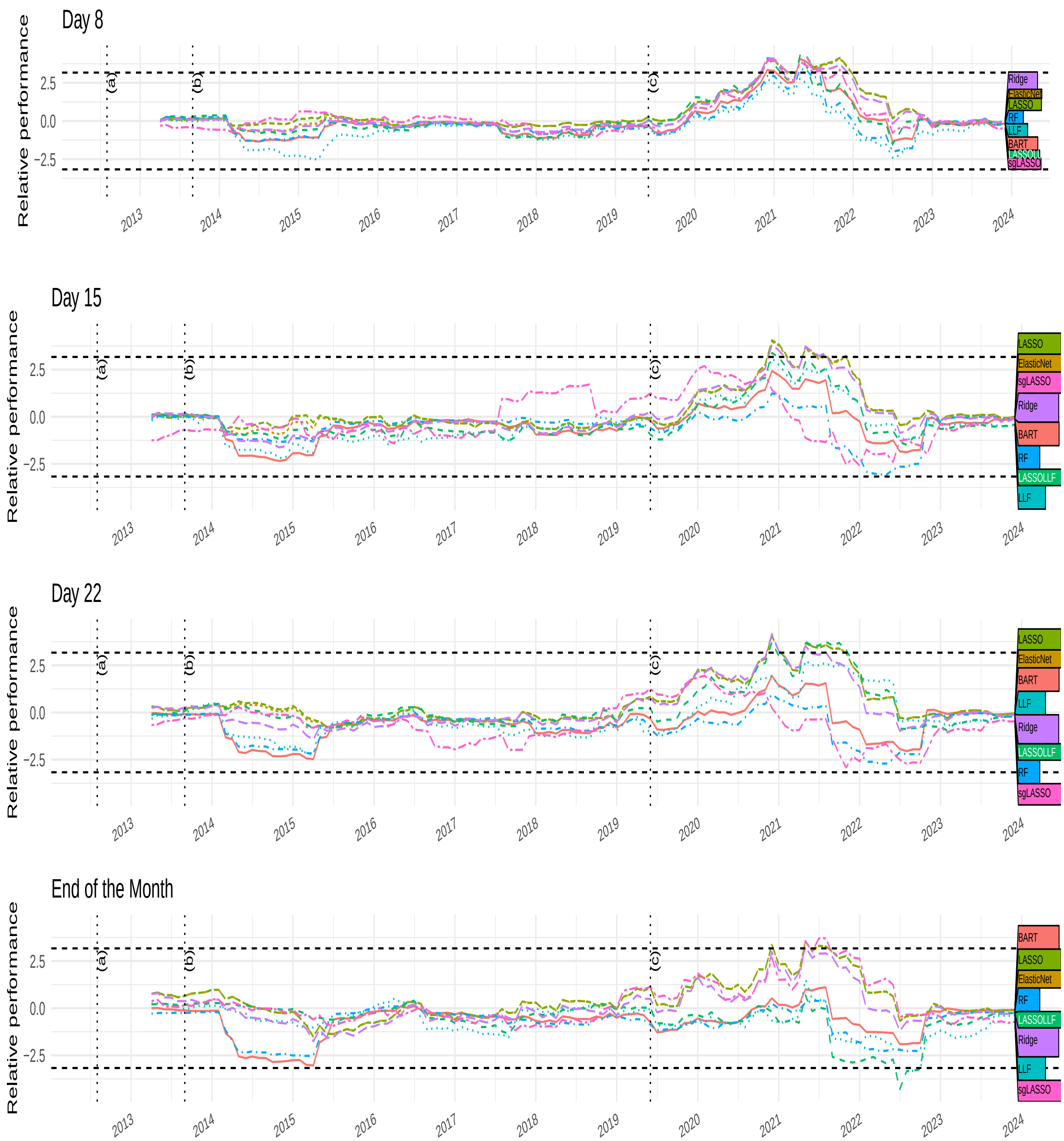
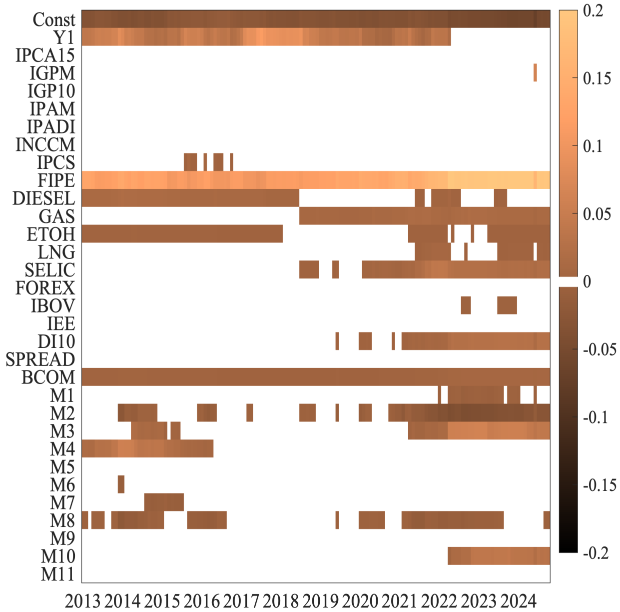


Figure B2: Fluctuation test: ML competing models versus the SPF benchmark

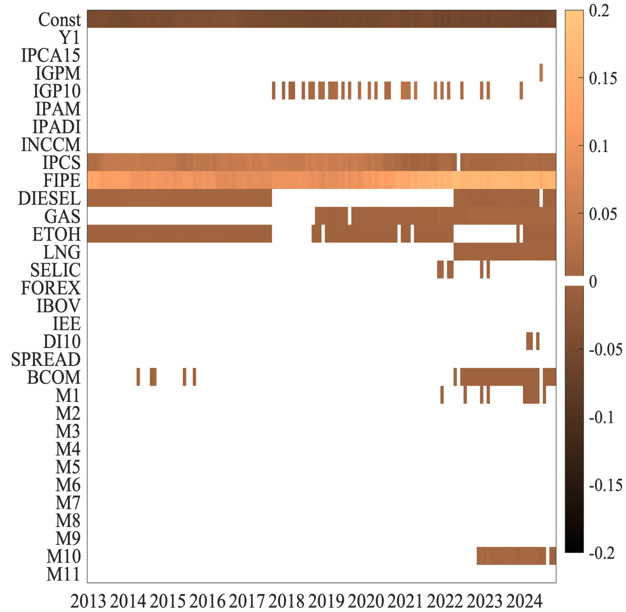


Notes: This Figure reports the fluctuation test from Giacomini and Rossi (2010) based on the squared loss differential between a machine learning method and SPF nowcasts. Areas between the horizontal dashed lines correspond to the 90% confidence interval of the two-sided statistical test. We used as window parameters of the test $\mu = 0.1$ and five for the number of lags in the variance of the DM test.

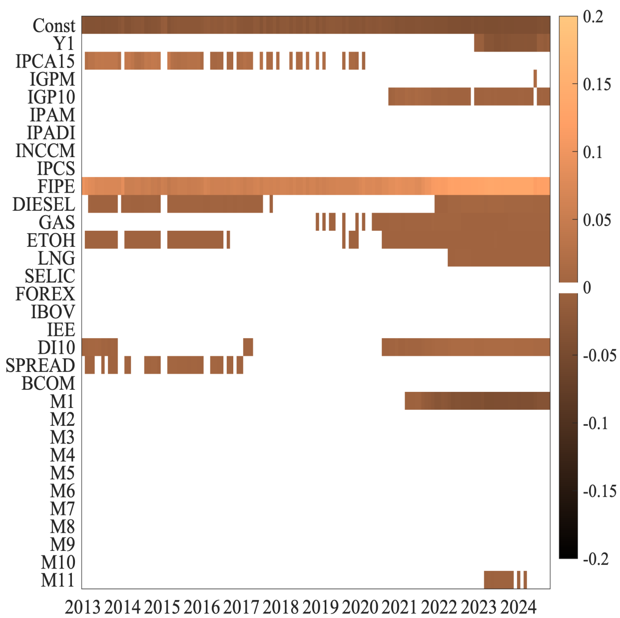
Figure B3: Heatmap of coefficient estimates using LASSO on the SPF nowcasting errors



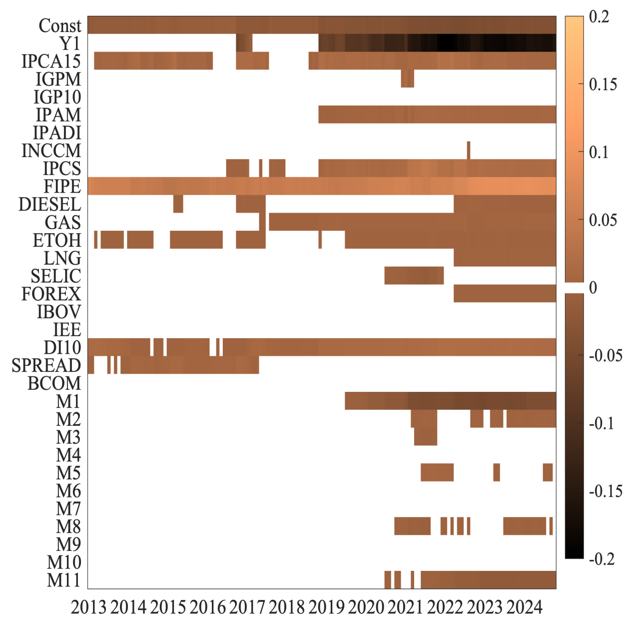
(a) Day 8



(b) Day 15



(c) Day 22



(d) End-of-month

Notes: This Figure depicts heatmaps of LASSO-fitted coefficients using SPF nowcasting errors as the dependent variable. Empty cells represent a coefficient estimate equal to zero, and thus a predictor that has not been selected at the estimation round t in the evaluation period.