Assessing forecasting in linear and non-linear Machine Learning models for short-term inflation in Peru

Esteban Cabrera Bonilla*
Pontificia Universidad Católica del Perú

This Version: May 17, 2024

Abstract

Inflation forecasting is a key element of monetary policy. A relatively recent approach to forecasting has been the implementation of Machine Learning methods. In this document we compare two types of methods, benchmark models, commonly used to predict future inflation, with linear and non-linear Machine Learning methods for inflation forecasting in Peru after the implementation of an inflation-targeting regime. We test the models applying a twelve-horizon forecast in five different periods (from 2019 to 2023). We demonstrate that both linear and non-linear machine learning models can easily outperform traditional models in an emerging economy with a steady price level like Peru, especially after the pandemic. We find out non-linear models are most suitable for inflation forecasting in the presence of high volatility in inflation, particulary, during the pandemic and its aftermath. This might imply the presence of non-linear relationships between inflation and its predictors during periods of external shocks and high volatility.

JEL Classification : E31, C49, C53, C59.

Keywords: Inflation, forecasting, Machine Learning, Peru.

*E-Mail Address: esteban.cabrera@pucp.edu.pe.

1 Introduction

As Stock and Watson (2001) said, macroeconometricians are given four main tasks:

> *"describe and summarize macroeconomic data, make macroeconomic forecasts, quantify what we do or do not know about the true structure of the macroeconomy, and advise macroeconomic policymakers."*

In this thesis, we will focus on the second and fourth tasks, macroeconomic forecasts and policy advice. Inflation is one of the most important macroeconomic indicators for a country, therefore it is crucial to keep it monitored. Across the world, one thing that most central banks have in common is the goal to keep the steadiness of price levels. In Peru, for instance, the main purpose of the Central Bank of Reserve (BCRP) is "to preserve monetary stability" (BCRP, n/d), which translates into keeping inflation under control. Similarly, monetary authorities of other countries in the region pursue the same objective.

Inflation forecasting is key to dictating an accurate monetary policy. Due to the lag between implementing monetary policy and its effects, most central banks implement a forecast-based monetary policy. In that sense, inflation forecasting becomes one of the major tasks of policymakers since it is one of the main indicators that will drive monetary policy, the primary tool to target inflation. Forecasting also contributes to macroeconomic stabilization, as it helps private agents make better and more informed decisions (Carnot et al., 2011).

In 2001, the BCRP adopted the reference interest rate as its main policy instrument. The following year, 2002, it implemented an inflation targeting system to make the monetary policy decisions more transparent. These two changes, alongside other measures, such as the accumulation of large foreign exchange reserves on a large scale, changed the dynamics of inflation in Peru and allowed the central bank to ensured monetary stability in Peru in favourable and unfavourable external contexts over the first decades of the 21st century (Dancourt, 2014). When it comes to forecasting, these changes have made Peruvian inflation easier to forecast as it has become more stable since volatility has decreased.

The BCRP uses the dynamic semi-structural model known as *Modelo de Proyección Trimestral* (MPT) as its main forecasting and policymaking tool. However, it has been pointed out that the model lacks parsimony, as it is composed by hundreds of equations. Llosa et al. (2006) proposes, alternatively, the use of Bayesian techniques and a small number of variables to accurately forecast Peruvian inflation. Barrera (2007) shows a robust sparse model approach to achieve parsimony, as these models identify the most relevant variables and discard the rest, reducing the number of variables used in the forecast.

Machine Learning (ML) is a relatively new approach to inflation forecasting. Therefore, there is little literature on the subject compared to other forecasting methods. Nevertheless, many of the papers on ML have concluded that both linear and non-linear multivariate ML models outperform other forecasting methods under particular scenarios (Ülke et al., 2018; Medeiros et al., 2022). In particular, they find out that non-linear models, like Random Forest, are way better at predicting inflation than most other models. The recent increase in volatility of inflation, due to shocks of different nature, such as the pandemic in 2020, the war in Ukraine and its subsequent increase in commodity prices, among others, make room for the introduction of these new methods.

Across the region, there have been articles that have compared traditional univariate and multivariate econometric models with ML methods in terms of forecasting (Rodriguez-Vargas, 2020;

Silva & Piazza, 2020). However, there has not been a proper comparison against more potent econometric methods. Furthermore, such an analysis with ML models has been just been recently done in Peru (Flores & Grandez, 2023), leaving room to expand the research in this topic.

Therefore, the purpose of this thesis is to evaluate the performance of different linear and non-linear univariate and multivariate models in terms of forecasting inflation for Peru after the introduction of the explicit inflation-targeting regime. Our benchmark econometric model will be a Random Walk (RW). We will compare econometric models such as a Vector Autoregression (VAR), while the ML methods will be a LASSO regression, ridge regression, elastic net (EN), random forest (RF) and support vector regression (SVM). We will see which models perform better during five main forecasting periods within a twelve-month forecast horizon, January to December, spanning the years 2019 to 2023. We will see which variables contribute more to making better predictions. We will also find out if non-linear models are most suitable for inflation forecasting in Peru, which would imply the presence of non-linear relationships between inflation and its predictors during periods of high inflation.

The structure of the document includes the following: Section 2 outlines both the theoretical and empirical literature regarding inflation forecasting; Section **??** describes the theoretical framework that is going to be used to compare the forecasting models; Section 3 portrays the data and describes the models; Section 4 comprises the discussion of the main results, followed by the concluding section.[1]

2 Literature review

Contrary to appearances, ML has an extensive history in macroeconometrics. Early works of Lee et al. (1993) and Kuan and White (1994) show the aplication of neural networks and other ML methods to macroeconomic series. When it comes to forecasting, Stock and Watson (1999) summarize the results of time series forecasting with multiple linear and nonlinear models with different specifications, with a total of 49 different methods, including simple autoregressions, exponential smoothing, artificial neural networks (ANN) and logistic smooth transition autoregressions (LSTAR). They conclude that non-linear methods like ANN and LSTAR operate optimizing a local optima, which do not necessarily translate into a global optima. Furthermore, improving an in-sample loss function does not result in an accurate out-of-sample forecast. Therefore, non-linear models are easily be surpassed by the autoregressions. Swanson and White (1997) compare multivariate ANN models with vector autoregressions, similarly concluding that the latter ones have lower mean square errors (MSE). On the other hand, Nakamura (2005) compares ANN with autoregressions in a pseudo-out-of-sample forecasting with US data, finding that ANNs outperform autoregressions, especially in short horizons.

The application of ML to inflation forecasting has focused on developed countries. For instance, Ülke et al. (2018) make inflation predictions using both time series and ML for the USA with data between 1984 and 2014. The study shows that ML prevails against the time series models in at least seven out of the sixteen conditions used.

Recently, ML literature has popularized among developing countries. In India, Pratap & Sengupta (2019) compare the inflation series forecast of ML algorithms with standard statistical models, finding that ML methods are generally able to outperform standard models. Pradhan (2011) uses ANN and univariate models to forecast Indian inflation, concluding that multivariate ANN models

[1] These are preliminary conclusions.

outperform univariate models in mean squared errors and mean absolute deviation. Özgür & Akkoç (2021) compare ML models with univariate models in the forecast of Turkish inflation, arriving to the conclusion that the former models reduce errors more.

Medeiros et al. (2022) contributed with a more detailed investigation that included a total of 91 countries, with observations from January 1980 to December 2019. They compared six models, three traditional models, and three ML models, concluding that, overall, ML models, particularly non-linear ones, were the best suited for forecasting inflation. Although their research was conducted for a large group of countries, the results were the same for developing economies like Brazil and Nigeria, where the RF model had the best metrics, especially in horizons from three to six months.

One article centred around ML forecasting for a country in the region was done by Silva & Piazza (2020). Their objective was to build accurate forecasts of the Brazilian consumer price inflation (IPCA) at multiple horizons spanning from one to twelve months (h = 1, ..., 12). The used the survey forecast as a benchmark. In the shortest horizon (h=1), the best model was the iterated factor model, followed by the survey forecast. In the third and fourth places were a reduced-form structural model (hybrid Phillips curve model) and the VAR (34) model. For longer horizons, the survey forecasts still outperformed other models in terms of mean squared error (MSE), although very often with equal predictive ability to the factor model and the random forest models.

Medeiros et al. (2016) also published a work on inflation forecasting using ML for Brazil. Specifically, the variable to forecast was the National Consumer Price Index (IPCA) and the monthly inflation index (IGPM). The results showed that adaLASSO was the best model to forecast the IPCA inflation in the horizon from one to four months (h = 1, , 4). On the other hand, for longer horizons (h = 5, , 12), the AR and the factor models had better predictions. Regarding the IGPM inflation, the adaLASSO was again the best model for short-period forecasts, but the AR model was best suited for longer horizons.

Similarly, Rodriguez-Vargas (2020) published another work focused on a developing country. The author chose the interannual variation rate of the Consumer Price Index of Costa Rica as the variable to forecast for one, three, six, and twelve months (h = 1, 3, 6, 12). Using as metrics of comparion the MSE and the Theil index, the results showed that at all horizons, the LSTM model produces the most accurate predictions, followed by the average of the univariate methods, the forecasts of the univariate KNN, and those of extreme gradient boosting and random forests. He also concluded that a combination of forecasts can improve the performance in comparison with individual forecasts at all horizons, and, most importantly, outperforms the forecasts from univariate methods.

In Peru, there is also literature that has used these innovative approaches to forecast inflation. Barrera (2007) uses an Unaggregated Forecasting System (Sistema de Predicción Desagregada or SPD) consisting in a Sparse VAR model approach to predict Metropolitan Lima CPI. A Sparse VAR model can handle a huge number of equations while maintaining parsimony in the model by giving emphasis on identifying and modeling only the most relevant relationships while ignoring or assigning low weights to the rest of the relationships between variables. Sparse VAR parameters are usually sensible to outliers. However, Barrera proposes a robust multi-equation procedure that presents a gain in accuracy in the presence of outliers. The author concludes that robust SparseVAR models improve the accuracy of all vanilla SparseVAR models for intermediate horizons, especially in three of the four types of SparseVAR models. Robust projections are less sensitive to outlier data sequences such as those experienced during the 1998 ENSO phenomenon and are thus suitably designed for the eventual occurrence of this phenomenon in the future.

Literature on inflation forecasting using ML for Peru is quite recent, with the work of Flores & Grandez (2023) being one of the first. The authors compare nine different ML models against two traditional benchmarks. Their selected period covers from 2002 to 2023. The authors conclude that LASSO and Random Forest are competitive in the short and long term, respectively, and that the combination of forecasts (simple and weighted) outperforms across most prediction horizon.

3 Methodology

In this section we present the methodology to assess the models' performance. Models are ranked based on their RMSE and their MAPE for each period in all three horizons. We then use the Diebold-Mariano test to compare each model against the benchmark RW and see how each model competes it. We are considering the two econometric models (RW, VAR) and six machine learning models (LASSO, Ridge, LARS, EN, RF and SVR).

3.1 Data

Our analysis employs

Our analysis employs nine distinct series, notably featuring the Metropolitan Lima Price Index represented in monthly percentage change over a twelve-month period (CPI) as our outcome variable. Our main predictors are national variables such as the monetary policy rate, the exchange rate, the money supply (M1), the international net reserves and the real minimum wage index. Furthermore, our model integrates external variables such as Maize (US$ per tonne), Wheat (US$ per tonne), Soybean oil (US$ per tonne), and Crude Oil (US$ per barrel). All data utilized in this study is sourced from the database of the Central Reserve Bank of Peru (BCRP), with a consistent monthly frequency.

The graphical representations in Figure A1 and A2 depict the plotted trajectories of these variables over time. Notably, the graphs reveal three distinct periods characterized by heightened volatility: the periods of 2007-2008, 2014-2015, and most recently, 2020-2022. For the purpose of our forecasting analysis, we are focusing only in the most recent period of high volatility. The idea is to compare the performance of all the models under different circumstances. First, a low-volatility period, during 2019. In second and third place, we will see how the models behaved during the pandemic shock and in the subsequent recovery, which was characterized by a huge increase in volatility. Finally, in the fourth and fifth forecast we look at the performance of the models in recent years, characterized by external shocks, as well as a strong central bank response.

3.2 General Framework

Given the monthly price level P_t, we define the monthly inflation as $\pi_t = 100 \times (ln(P_t) - ln(P_{t-1}))$. Let's assume a $K \times 1$ vector X_t of predictors. Our objective is to predict inflation h periods forwards π_{t+h}, which can we viewed as:

$$\pi_{t+h} = F_h(X_t, \gamma) + \epsilon_{t+h}$$

where $h = 1, ..., H$ is the forecast horizon. $F(.)$ is the relationship between inflation and its predictors, which could either linear or non-linear, depending on the model being used, γ represents both parameters and hyperpameters of the ML models and ϵ_{t+h} is the forecast error.

3.2.1 Forecasting Procedure

To perform the analysis of the predictive power of the different ML models, we first standardize the data to ensure all features are on the same scale. We divide our data into two consecutive sub-samples: training and testing. In the training sample we will both fit the model and calibrate the hyperparameters. To do that we implement a time series cross-validation, which, unlike other forms of cross-validation, considers the structure of the time series. The data is split into k-folds. Each iteration, the j-fold is used as the validation set, and the remaining $j-1$ folds are used as the training set. The optimization parameters in γ are choose to minimize a metric, in this case the mean square error. The process is repeated until all folds have been used to calibrate the tuning parameters. The final hyperpameters are those that on average minimized the metrics across the different folds. This means that the ML model is trained a total of k times. Once the model has been calibrated, we use it to do out-of-sample forecasting in the testing sample. Ultimately, we rank the models based in their performance in the out-of-sample forecast.

3.3 Forecast Evaluation

The most common approach when ranking forecasts is to compare their accuracy. That is, the measure of how accurately a given forecast matches actual values. In this subsection we first portray deterministic comparison methods such as the root mean squared forecast error (RMSFE) and the mean absolute percentage error (MAPE). Then, following Ghysels & Marcellino (2018) we outline a forecast comparison test that evaluates whether the difference between forecasts is statistically significant, namely the Diebold-Mariano test.

3.3.1 Root Mean Square Forecast Error (RMSFE)

Given the forecast error

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h}$$

The MSFE can be depicted as

$$MSFE = \frac{1}{T}\sum_{t=1}^{T}(y_{t+h} - \hat{y}_{t+h})^2$$

$$MSFE = \frac{1}{T}\sum_{t=1}^{N} e_{t+h}^2$$

Therefore, the RMSFE is defined by

$$RMSFE = \sqrt{MSFE} = \sqrt{\frac{1}{T}\sum_{t=1}^{N} e_{t+h}^2} \quad (1)$$

3.3.2 Mean absolute percentage error (MAPE)

The mean absolute percentage error is a loss function defined as

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{t+h} - \hat{y}_{t+h}}{y_{t+h}}\right|$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}|\frac{e_{t+h}}{y_{t+h}}| \tag{2}$$

3.3.3 Diebold-Mariano test

Proposed by Diebold & Mariano (1995), this test relaxes the requirement of forecasts errors and can compare directly general loss functions. Given the forecast errors e_i of the competing forecast models i for $i = 1, 2$, we can define the loss-differential as

$$d_j = C(e_1 j) - C(e_2 j)$$

where C is the loss function, which, for example, can be the quadratic loss $C(e) = e^2$ or the absolute loss$C(e) = |e|$. We want to test the null hypothesis

$$H_0 : E(d_j) = 0$$

against the alternative

$$H_1 : E(d_j) \neq 0$$

Then, the Diebold-Mariano statistic is defined as

$$DM = H^{1/2}\frac{\sum_{j=1}^{H} d_j/H}{\sigma_d} = H^{1/2}\frac{\bar{d}}{\sigma_d} \tag{3}$$

σ_d is the variance of $\hat{d}$, which can be estimated given that

$$\hat{\sigma_d} = (\gamma_0 + 2\sum_{i=1}^{h-1}\gamma_i)$$

$$\gamma_k = H^{-1}\sum_{t=k-1}^{H}(d_t - \bar{d})d_{t-k} - \bar{d})$$

where h is the forecast horizon, meaning that in $h = 1$ there is no correlation, so the standard formula for the estimation of the variance γ_0 can be used.

3.4 Econometric models

In this subsection we will describe two frequentist approaches to inflation forecasting: an univariate random walk and a multivariate vector autoregression.

3.4.1 Random Walk (RW)

Out of all the models, the RW is simplest. Consider the non-stationary model

$$y_t = y_{t-1} + \epsilon_t \tag{4}$$

For any h periods ahead being forecasted it assumes that the inflation rate is predicted with the last observation of itself.

$$\hat{y}_{T+h} = y_T \tag{5}$$

The forecast errors are

$$e_{T+h} = \epsilon_{t+1} + \epsilon_{t+2} + ... + \epsilon_{t+h}$$

Hence, the variance of the forecast errors are

$$Var(e_{T+h}) = h\sigma_\epsilon^2$$

3.4.2 Vector Autoregression (VAR)

Given a VAR(p) model of k variables$y_{1t}, ..., y_{kt}$ grouped in a $(k \times 1)$ vector $y_t = (y_1 t, ..., y_k t)'$, it can be represented as:

$$y_t = \mu + \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + \epsilon_t \tag{6}$$

$$\epsilon_t \sim WN(0, \Sigma) \tag{7}$$

where Σ is the $(k \times k)$ variance-covariance matrix and

$$\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{kt})'$$

is a vector of error terms. Each error is uncorrelated over time and homoskcedastic, but can be correlated with errors in other equations, meaningϵ_t is a multivariate white noise process.

Note that the total number of parameters grows very fast as the number of variables k increases. There are k parameters in the intercept, $k^2 p$ coefficients of the lagged variables and $k(k+1)/2$ parameters in the variance-covariance matrix.

Using the Wold decomposition theorem, Equation (12) can be represented as VMA(∞)

$$y_t = \Psi(L)\epsilon_t \tag{8}$$

where $\Psi(L) = (I + \Psi_1 L + \Psi_2 L^2 + ...)$ is a matrix polynomial in L.

The optimal point forecast for the VAR(p) is the extension of the univariate case, meaning that

$$\hat{y}_{T+h} = \Phi_1 \hat{y}_{T+h-1} + ... + \Phi_p \hat{y}_{T+h-p}$$

Using the VMA(∞) representation in (17). The optimal forecast can be rewritten as

$$\hat{y}_{T+h} = \sum_{j=0}^{\infty} \Psi_{j+h} \epsilon_{T-j} \tag{9}$$

with the following forecast error

$$e_{t+h} = \sum_{j=0}^{h-1} \Psi_j \epsilon_{t+H-j} \tag{10}$$

and the variance-covariance of the forecast error

$$V(e_{t+h}) = \Sigma + \Psi_1 \Sigma \Psi_1' + ... + \Psi_{h-1} \Sigma \Psi_{h-1}' \tag{11}$$

3.5 Machine Learning models

In this subsection we will describe the ML models. We will describe two linear methods, namely LASSO regression and Elastic Net, and one non-linear method: Random Forest.

3.5.1 LASSO Regression

The least absolute shrinkage and selection operator (LASSO) was first developed as a frequentist shrinkage method by Tibshirani (1996). In ML it is used as method for feature selection and regularization. The LASSO regression adds a penalty term which depends on the absolute value of the regression coefficients.

Given the following multivariate linear regression model

$$Y = X\beta + \epsilon$$

where Y is an $N \times 1$ vector of dependant variables, X is an $N \times K$ matrix of explanatory variables, $\beta = (\beta_1, ..., \beta_k)$ is a $K \times 1$ vector regression coefficients and ϵ is an $N \times 1$ vector of errors. It is possible that K is relatively large compared to N. In those cases, the LASSO estimates are chosen to minimize

$$LASSO = \min_{\beta} \left\{ (y - X\beta)^T (y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{12}$$

$$LASSO = \min_{\beta} \left\{ \|y - X\beta\|^2 + \lambda \|\beta\|_1 \right\} \tag{13}$$

where $\|y - X\beta\|^2$is the residual sum of squares (RSS) and the second term $\lambda\|\beta\|_1$ is the ℓ_1 regularization penalty term, also known as the shrinkage penalty. It is null when$\beta_1, ..., \beta_k$ are set to zero. $\lambda \geq 0$ is a tuning parameter, it is use to control the impact of the coefficients in the regression. When $\lambda = 0$ there is no penalty effect. The greater λ is, the bigger the penalty, that means that more coefficients β_i will be equaled to zero. LASSO yields sparse models, which are models that involve only a subset of the variables. This process of reducing variables is called feature selection. Upon completion, we proceed to do the point forecasts.

3.5.2 Ridge Regression

Consider the same linear regression model as in LASSO, the ridge coefficients are chosen by imposing a penalty on the squared estimates:

$$Ridge = \min_{\beta} \left\{ \|y - X\beta\|^2 + \lambda \|\beta\|_2^2 \right\} \tag{14}$$

where the term $\lambda\|\beta\|_2^2$ is a regularization of type ℓ_2 and λ is the tuning parameter.

As λ increases, the coefficients $\beta_1, ..., \beta_k$ will approach zero. Notice that, unlike the LASSO regression, the estimates cannot be zero. That means that given a model with a large number of parameters, the ridge regression will always generate a model involving all predictors, but will reduce their magnitudes by making the coefficients are very small. After the estimation of the the ℓ_2 penalty, we proceed with the point forecasts.

3.5.3 Least Angle Regression

Least Angle Regression (LARS) is an alternative method for feature selection and regularization introduced by Efron et al. (2004). Similarly to other regularization methods, it is particularly useful with high-dimensional data where the number of predictors is relatively large compared to the number of observations. This algorithm is also more computationally efficient compared with LASSO.

Initialization:

1. We start with all coefficients equal to zero and compute the initial residuals, which are just the observed values
$$\beta^{(0)} = 0, \quad \text{residual} = \mathbf{Y}$$

Iteration:

1. At each step k, find the predictor x_j most correlated with the current residuals. This involves finding the predictor x_j that has the highest absolute correlation with the residuals $\mathbf{Y} - \mathbf{X}\beta^{(k)}$. We first compute the correlation c_j between each predictor and the residuals and find the predictor with the highest absolute correlation.
$$c_j = \mathbf{X}_j^T(\mathbf{Y} - \mathbf{X}\beta^{(k)})$$
$$j = \arg\max_j |c_j|$$

2. Move the coefficient β_j in the direction that reduces the RSS (in the direction of the sign of c_j). Update the residuals as the coefficient changes. Continue this movement until another predictor x_k achieves the same absolute correlation with the residuals.
$$\beta_j^{(k+1)} = \beta_j^{(k)} + \gamma, \quad \text{residual} = \mathbf{Y} - \mathbf{X}\beta$$

3. Once another predictor $x_k$has the same correlation with the residuals, move the coefficients β_j and β_k in a direction that keeps the current residuals equally correlated with both predictors.
$$\beta_j^{(k+1)} = \beta_j^{(k)} + \gamma_j, \quad \beta_k^{(k+1)} = \beta_k^{(k)} + \gamma_k$$
$$|\mathbf{X}_j^T(\mathbf{Y} - \mathbf{X}\beta)| = |\mathbf{X}_k^T(\mathbf{Y} - \mathbf{X}\beta)|$$

4. Continue steps 1-3 until all predictors have been considered or until the desired number of predictors have been included in the model.

In the LARS model, the hyperparameter used is the number η of non-zero coefficients. This number will determine the desired number of predictors that are included. η is chosen to minimize the mean square error.

3.5.4 Elastic Net

Elastic Net is another regularization technique used in linear regressions. It combines both ℓ_1 and ℓ_2 regularization's. That means that the estimates will be chosen by

$$Elastic\ Net = \min_\beta \left\{\|y - X\beta\|_2^2 + \lambda\left(\rho\|\beta\|_1 + (1-\rho)\|\beta\|_2^2\right)\right\} \tag{15}$$

where $\|\beta\|_1$ and $\|\beta\|_2^2$ correspond to their specific regularization and ρ is a tuning parameter that will measure the weight of theℓ_1 penalty. If $\alpha = 0$, then we will have a Ridge regression. If it is 1, then the EN transforms into a LASSO regression. By convention, we use $\rho = 0.5$ which will give an equal proportion to each regularization technique. The other tuning parameter λ presents the weight of the combined penalties. It is chosen during the cross-validation step to minimize the mean square error.

3.5.5 Random Forest

RF is a non-linear technique that can be used to create a large number of regression trees. This regressions trees divide the observations into regions where the predictor values are similar. The choice of the splitting points is done by minimizing a loss function. In this way the overall variance is reduce by averaging many of these regression trees. To understand the concept better, let us first compare the typical linear regression model

$$y = \sum_{i=1}^{p} X_i \beta_i \tag{16}$$

where there is linearity in the coefficients. That means that the relationship between the dependant and the independent variables is a linear combination of x and β . Now let us assume a regression tree model in the form

$$y = \sum_{m=1}^{M} c_m \cdot 1_{(x \in R_m)} \tag{17}$$

where $(R_1, ..., R_M)$ correspond to the partition regions for the observations. To construct the regression tree, a set of possible values $(x_1, ..., x_p)$ is split into M possible non-overlapping regions $(R_1, ..., R_M)$. Then, for every observation that falls into the region R_m, we will make the same prediction, which is the average of the response values for the training observations in R_m.

In the context of RF, each time a split in a tree is done, a random sample of m predictors is chosen as split possible candidates from the full set of M predictors. This split is allowed to use only one of those m predictors. Usually, the number of predictors assessed at each split is approximately the square root of the total number of predictors M, meaning $m = \sqrt{M}$, which differentiates RF from bootstrapping, where the split considers the full sample $m = M$ each time. The RF way of splitting predictors will typically be useful when we have a large number of correlated predictors in our dataset, which could be the case of inflation.

If the relationship between the predictor x and the response variable y is linear, then (18) will be the ideal model to use, as it will outperform the regression tree. However, in the presence of non-linearity's in the features and the response, then the model (19) will have better results. If there is a presence of non-linear relationships between inflation and its predictors during the analyzed period, then the RF model will outperform all linear econometric and ML models.

3.5.6 Support Vector Machine Regression (SVM)

4 Results

In this section we will present the main results after running the models. We are comparing both the RMSE and the MAPE through horizons 1 to 12 for five different periods. We are using the first to fourth lags of the other variables as predictors for headline inflation.

4.1 2019

The first period where the forecast models are being tested is 2019, from January to December. This year is characterized by a low volatility in headline inflation, our outcome, as well as the other input variables. We can see in figure A5 the different predictions of each of our five models, as well as the actual inflation plotted in orange. While most of the models capture the slight downfall of inflation during the following twelve months, they fail to catch the exact magnitude of it. The LASSO model, on the other hand, completely misses the forecast, predicting a straight line, just like the RW, but with higher values for the inflation.

In table A1 and A2 , we can see the RMSE and MAPE of the predictions of the benchmark, as well as the ratios of these indicators for the other models, with a number higher than one meaning the RW offered a better prediction, while a number lower than one indicates that the model surpassed the benchmark. We see that most models have high ratios in the first four horizons, both in terms of RMSE and MAPE, with the RF having the lowest ratio. It is closely followed by the non-linear RF. We observe that most models struggled in the shortest horizons, but the RF had the best results from $h = 1$ to $h = 4$. However, it is quickly surpassed in the long-term horizon by the traditional VAR model. From around $h = 6$ onwards, the VAR model outperforms the ML models, as well as the benchmark, having ratios lower than one.

In terms of predictors, it is interesting to analyze the variables chosen by both the linear and non-linear ML methods. This is shown in figure A6 and A7 . The ridge model chose the first lag of the interest rate, the soybean oil price, the net national reserves and the exchange rate, assigning them a positive coefficient. It also chose the fourth and third lag of the monetary policy rate, giving them a negative coefficient, capturing this negative lagged relation between inflation and the interest rate. The RF model, on the other hand, did not choose as main features any of the lags of the reference rate, but the lags of the wheat price per ton and the circulating currency. This could indicate that the monetary policy rate is more linearly related to headline inflation, as it was primarily chosen by the ridge model. The LASSO model, on the other hand, did not choose any feature, asigning a weight of zero to every predictor. This could explain its poor performance compared to the random walk.

4.2 2020

The second period is characterized by the pandemic, which greatly affected our predictors as well as our response variables. The interest rate fell as a response of the central bank to the crisis, there was also a huge but quick plunge in crude oil prices and a rise in many aggricultural commodity prices, as well as the circulating currency. The massive change caused by the the pandemic hindered the performance of the traditional econometric models, mainly the VAR. We can see in table A4 and A5 it had massive ratios in both RMSE and MAPE, compared to the benchmark, indicating a poor performance, especially in March, when the pandemic hit the hardest. We observe a similar behaviour for the linear ML models. Nevertheless, their performance improve over the following months, in the longer horizons. The best model in this pandemic-hit period is the RF, which had the lowest ratio and even outperformed the RW in the longest horizons ($h = 10$ to 12). The linear ML models also have a better performance than the VAR, mainly after $h = 6$, but they still leave much to be desired. In figure A8 we see the each forecast plotted. The RF model captured pretty well the surge in inflation. The linear models (ridge, LASSO) tracked it well from the third horizon onwards, while the VAR model completely missed the headline inflation behaviour.

Regarding the predictors selected by each ML model, which are shown in figures A9 to A11 , it is interesting that both the LASSO and ridge models algorithms only opted for the same three predictors: the first lag of the soybean oil price, the first lag of the policy rate and the second lag of the wheat price. The RF picked a wider set of predictors, including the multiple lags of the wheat price, the crude oil price and the money supply, the same features chosen in 2019.

4.3 2021

The third period corresponds to the post-pandemic recovery. It is distinguished by an increase in volatility in the exchange rate and other external variables, many of which reached their highest point halfway through the year and continue increasing the following months. There was also a response of the central bank to the inflation spike which greatly increased the interest rate. The recovery period also saw a downfall in the performance of the traditional VAR method, as it is surpassed by both linear and non-linear ML methods. Between the linear and non–linear methods, the former ones show an advantage, mainly the ridge algorithm, which has the lowest RMSE and MAPE from $h = 5$ until the end of the forecast, as we see in table A7 and A8. In figure A12 we can compare the different forecasts of each model againts the actual inflation (in orange). The VAR model again does a poorly job in predicting the movements of the the outcome variable.

In figure A13 to A15 we see the predictors selected by each model. This time, the ridge and LASSO algorithms had different selections, with the ridge method choosing a wider set of features. The LASSO model chose the same three features as in 2019. Meanwhile, the ridge model selected as its main predictor the first of the moneraty policy, followed by the lags of the external variables (wheat and soybean oil prices). It also captures a positive relation between the lags of the exchange rate and inflation, and a negative one between the third and fourth lag of the policy rate. The RF model was again the one that selected the most predictors, being the most important features (as indicated in figure A15) the lags the wheat price and the money supply, followed by other external variables.

4.4 2022

The fourth period, from January to December 2022, was marked by the continuous increase of the monetary policy, as well as a persistent high volatility in both inflation and the prices of the commodities. In figure A16, we can observe the forecasts of each different model. The VAR again shows a poor performance, compared with the ML models, predicting a rapid fall which did not occur. The linear models (LASSO and ridge) capture the fluctuations of inflation but missed the actual values. The RF model had a good performance in the near-term horizon. In terms of RMSE and MAPE, the best models are RF from $h = 1$ to $h = 5$ and ridge from $h = 6$ onwards.

The predictors elected by the ridge model are mainly the lags of commodities and the monetary policy rate, as shown in figure A17. The LASSO model . The RF technique picked the lags of commodities, as well as the exchange rate and the currency in circulation, which is shown in figure A19.

4.5 2023

The final period is characterized by more stability than the previous periods. There was less volatility of the commodity prices and there was a moderate decrease in the interest rate in the latter months. In figure A20, we see that all the ML methods, both linear and non-linear, catpured well the behaviour. The VAR model initially has a poor performance but it improves by the end. In table A13 and A14, we can see the RMSE and MAPE of the different models as a ratio of the RW values. All ML models have low values (close to one) in the shor-term horizon and improve upon the benchmark in the distant horizon. The ridge algorith has overall the lowest RMSE and MAPE quickly followed by the LASSO and RF models.

The predictors chosen by the ridge model are again the lags of the prices of some commodities (mainly soybean oil and wheat), the exchange rate and the reference rate. The last features picked by this linear model are the lag of the money supply. The LASSO model only selected ten features, being the most import the first lags of the soybean oil price, wheat price and exchange rate. The RF chose between its most important features the price lags of the price of soybean oil and wheat, as well as the lags of the exchange rate and the circulating currency.

4.6 Robustness

As a robustness exercise, we have performed the Diebold-Mariano test to evaluate if the differences between the forecasts of the different models and the RW are statistically significant. The results of the mean of the DM as well as the p-value can be found in the annexes for each corresponding year.

In 2019, we can see in table A3, despite being signs that the RW outperforms all models across the twelve horizons, there is no statistically significant difference between the RW and the VAR, ridge and RF models. Only the difference between LASSO and RW is statistically significant, with the latter one outperforming the former. In 2020, we see in table A6 that the only model the RW outperforms with statistically significance is the traditional VAR model. There is no indication that the RW outperforms any of the ML models. The results for 2021 indicate that no model performs better or worse than the RW. While there is some indication that the RF outperforms the benchmark, the p-value does not reach the 5% threshold. The 2022 results are more conclusive, as it is shown in table A12 that the benchmark outperforms each of the other more complex models, as it is indicated by the low p-value. Finally, in 2023, the RW appears to be better than the VAR, but it is not staticsally significant, and the RW appers to have worse forecast predictions than the ML models, as shown in table A15 but again, it is not statistically significant.

5 Conclusions

Our results show some unique findings regarding Peruvian inflation. First of all, it is noticed that is is hard for complex models to outperform the becnhmark RW when it comes to inflation forecasting. This is especially true for 2019, which was a year characterized by low volatility and a stable inflation. While the RF model outperform the VAR and he linear ML models in the short-horizon, the VAR model had the best results for the long-term forecasts. Nevertheless, our ML models were able to outperform both the VAR and the benchmark in most of the forecasts from 2020 onwards, having better results, in both RMSE and MAPE in the periods with higher volatility than usual. In particular, the non-linear RF model has lower metrics for the year 2020, which was the pandemic-hit year. In 2021, both ridge and LASSO linear models outperform the rest of the methods. In 2022 the RF and ridge models outperform the rest in short-term and long-term forecasting, respectively. In 2023, the linear ridge model has the best metrics for all horizons, followed closely by the LASSO and RF models.

Regarding the selection of predictors, it is interesting that most models selected the same variables across different years. The linear ML models, ridge and LASSO, consistently selected the lags of the monetary policy rate as one of their main predictors, as well as the price of soybean oil. The RF methods, on the other hand, did not select the interest rate among its main predictors, but instead choose the lags of the commodity prices, mainly wheat and soybean oil, as well as the circulating money and the exchange rate in some cases. It is also noticiable, that both linear models chose the same predictors in one of the forecast evaluation periods, during the pandemic. This could be an indicator of the type of relationship that exists among headline inflation and its predictors, with the selection of variables by the RF model possibly explaining some non-linearities. However, more research needs to be conducted on this matter. The SHAP values of the variables could also be calculated in order to better understand the relationship between our response and its predictors.

Our conclusions are similar to those presented in other studies and are in line with the recent work of Flores & Grandez (2023), showing that both the RF regression and linear ML models (ridge in our case, LASSO in theirs) have better results than other ML methods. This research should be expanded by adding more models and strengthening the robustness exercises.

6 Agenda

Esta sección esta en español. Se indicará la agenda de la investigación. Se deben agregar más modelos, tanto de ML como tradicionales. Basándonos en Flores & Grandez (2023), se debe agregar Elastic Net como modelo lineal, Decision tree como modelo no lineal y XGBoost y Gradient Boost como modelos ensemble, los cuáles también son no lineales. También se puede agregar la combinación de modelos, como hacen estos autores mediante un promedio simple y con pesos. Adicionalmente, se podría agregar las expectativas de inflación como un survey forecast, similar a Silva & Piazza (2020), para comparar a las expectativas con nuestros modelos. Se pueden agregar más variables, para mejorar el tunning de los distintos modelos, y evaluar agregar modelos univariados.

Además de ello, siguiendo a Medeiros et al. (2023), se podría computar los SHAP values, para poder interpretar mejor la elección de variables para los distintos modelos.

Finalmente, se deben reforzar los ejercicios de robustez, en particular, el Diebold-Mariano test.

References

[1] Anderson, O. D. (1977). Time Series Analysis and forecasting: Another look at the box-jenkins approach. *Journal of the Royal Statistical Society. Series D (The Statistician), 26(4)*, 285–303. https://doi.org/10.2307/2987813

[2] Banco Central de la República del Perú. (n.d.). Sobre el BCRP. Retrieved September 9, 2023, from https://www.bcrp.gob.pe/docs/sobre-el-bcrp/folleto/folleto-institucional-00.pdf.

[3] Banco Central de la República del Perú. (2015). (rep.). *Inflation Report. Recent trends and macroeconomic forecasts 2015-2017.* Retrieved October 3, 2023, from https://www.bcrp.gob.pe/eng-docs/Monetary-Policy/Inflation-Report/2015/inflation-report-may-2015.pdf.

[4] Bos, C. S., Franses, P. H., \& Ooms, M. (2001). Inflation, forecast intervals and long memory regression models. \textit{International Journal of Forecasting, 18(2)}, 243–264. https://doi.org/10.1016/s0169-2070(01)00156-x

[5] Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis, forecasting and control.* Holden-Day.

[6] Brown, T. M. (1954). Standard errors of forecast of a complete econometric model. *Econometrica, 22*, 178-192

[7] Carnot, N., Koen, V., & Tissot, B. (2011). Economic forecasting and policy (2nd ed.). Palgrave Macmillan.

[8] Clements, M. P., & Hendry, D. F. (1998). Forecasting economic time series. Cambridge University Press.

[9] Chan, J. C. C. (2013). Moving average stochastic volatility models with application to inflation forecast. \textit{Journal of Econometrics, 176(2)}, 162–172. https://doi.org/10.1016/j.jeconom.2013.05.003.

[10] Dancourt, O. (2014). (working paper). *Inflation Targeting in Peru: The Reasons for the Success* (pp. 1–55). Lima: Departamento de Economía.

[11] Doan, T., Litterman, R., & Sims, C. (1983). *Forecasting and conditional projection using realistic prior distrib* (Working Paper). National Bureau of Economic Research. https://doi.org/10.3386/w1202

[12] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals Of Statistics, 32(2).* https://doi.org/10.1214/009053604000000067

[13] Flores, J., & Grandez, R. (2023). *Proyectando la Inflación de Perú con Métodos de Machine Learning. Encuentro de Economistas.* Lima; Universidad del Pacífico. https://www.bcrp.gob.pe/docs/Proyeccion-Institucional/Encuentro-de-Economistas/2023/ee-2023-s20-p41-flores-grandez.pdf

[14] Ghysels, E., & Marcellino, M. (2018). Applied Economic Forecasting using time series methods. Oxford University Press.

[15] Granger, C. W. J. & Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics, 5*, 35-47.

[16] Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). Academic Press.

[17] Haavelmo, T. (1944). The Probability Approach in Econometrics. *Econometrica, 12*, iii. https://doi.org/10.2307/1906935

[18] Heesterman, A. R. G. (1970). Forecasting models for national economic planning. Springer.

[19] Heeren, S. (2020). *Forecasting Inflation using Machine Learning for an Emerging Economy* (thesis). Erasmus University Rotterdam.

[20] International Monetary Fund. (2008). *World Economic Outlook*. Retrieved October 3, 2023, from https://www.imf.org/en/Publications/WEO/Issues/2016/12/31/Global-slowdown-and-rising-inflation.

[21] Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921-41*. Cowles Comission Monograph. New York: John Wiley

[22] Lee, T. H., White, H., and Granger, C. W. J. (1993). Testing for neglected nonlinearity in time series models. A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56(3):269–290.

[23] Kuan, C. M. and White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Reviews*, 13(1).

[24] Litterman, R. B. (1980). *A Bayesian Procedure for Forecasting with Vector Auto-Regression.* (Working Paper) Massachusetts Institute of Technology, Department of Economics.

[25] Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions five years of experience. *Journal of Business & Economic Statistics, 4(1)*, 25–38. https://doi.org/10.21034/wp.274

[26] Llosa, G., Tuesta, V., & Vega, M. (2006). Un modelo de proyección BVAR para la inflación peruana. *Revista Estudios Económicos, (13).*

[27] Maravall, A. (1994). Use and misuse of unobserved components in economic forecasting. *Journal of Forecasting, 13(2)*, 157–178. https://doi.org/10.1002/for.3980130209

[28] Martin, V., Hurn, S., & Harris, D. (2013). Latent Factor Models. In *Econometric modelling with time series: specification, estimation and testing* (2nd ed., pp. 544–580). Cambridge University Press.

[29] Marget, A. W. (1929). Morgenstern on the methodology of economic forecasting.*Journal of Political Economy, 37*, 312-339.

[30] Medeiros, M., Schütte, E. C., & Soussi, T. S. (2022). Global inflation forecasting: Benefits from Machine Learning Methods. *SSRN Electronic Journal.*https://doi.org/10.2139/ssrn.4145665

[31] Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (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

[32] Medeiros, M. C., Vasconcelos, G., & Freitas, E. (2016). Forecasting Brazilian inflation with high-dimensional models. *Brazilian Review of Econometrics, 36(2)*, 223–254. https://doi.org/10.12660/bre.v99n992016.52273

[33] Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters, 86(3)*, 373-378.

[34] Özgür, Ö., & Akkoç, U. (2021). Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms. *International Journal Of Emerging Markets, 17(8)*, 1889-1908. https://doi.org/10.1108/ijoem-05-2020-0577

[35] Pradhan, R. P. (2011). Forecasting Inflation in India. *International Journal Of Asian Business And Information Management, 2(2)*, 64-73. https://doi.org/10.4018/jabim.2011040105

[36] Pratap, B., & Sengupta, S. (2019). Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts? *Social Science Research Network.* https://doi.org/10.2139/ssrn.3852945

[37] Pincheira-Brown, P., Selaive, J., & Nolazco, J. L. (2019). Forecasting inflation in Latin America with core measures. *International Journal of Forecasting, 35(3)*, 1060–1071. https://doi.org/10.1016/j.ijforecast.2019.04.011

[38] Pindyck, R. S., & Rubinfeld, D. L. (1981). Econometric models and economic forecasts. McGraw-Hill.

[39] Rodríguez-Vargas, A. (2020). Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking, 1(1-4)*. https://doi.org/10.1016/j.latcb.2020.100012

[40] Shiller, R. J. (1973). A distributed lag estimator derived from smoothness priors. *Econometrica, 41(4)*, 775–788. https://doi.org/10.2307/1914096

[41] Silva, G., & Piazza, W. (2020). Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models. *Banco Central Do Brasil.* Retrieved February 24, 2023, from https://www.bcb.gov.br/pec/wps/ingl/wps561.pdf.

[42] Stein, C. (1960). Multiple Regressions. In *Contributions to Probability and Statistics: Essays in Honor of Har*
Standford Univeristy Press.

[43] Stock, J. H., & Watson, M. W. (1988). Variable Trends in Economic Time Series. *Journal of Economic Perspectives, 2(3),* 147–174. https://doi.org/10.1257/jep.2.3.147

[44] Stock, J. H., & Watson, M. W. (1989). New Indexes of Coincident and Leading Economic Indicators. *NBER Macroeconomics Annual, 4*, 351–394. https://doi.org/10.1086/654119

[45] Stock, J. H., & Watson, M. W. (1993). A procedure for predicting recessions with leading indicators: econometric issues and recent performance. In *New Business Cycle Indicators and Forecasting.* Chicago University Press.

[46] Stock, J. H. and Watson, M. W. (1999). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In Engle, R. F. and White, H., editors, *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, pages 1–44.

[47] Oxford University Press, Oxford.

[48] Stock, J. H., & Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives, 15(4)*, 101–115.

[49] Stock, J. H., & Watson, M. W. (2003). *A probability model of the coincident economic indicators. In Leading e* Cambridge University Press.

[50] Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking, 39*, 3–33. https://doi.org/10.1111/j.1538-4616.2007.00014.x

[51] Svensson, L. (2005). Monetary policy with judgment : forecast targeting. *International Journal of Central Banking, 1*, 1–54.

[52] Theil, H. (1961). *Economic forecast and policy* (2nd ed.). North-Holland.

[53] Tinbergen, J. (1939). *Statistical Testing of Business-Cycle Theories.* Geneva: League of Nations. Vol I:*A method and its application to Investment Activity*; Vol II: *Business Cycles in the United States of America, 1919-1932*.

[54] Ülke, V., Sahin, A., & Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing & Applications, 30(5)*, 1519–1527. https://doi-org.ezproxybib.pucp.edu.pe/10.1007/s00521-016-2766-x

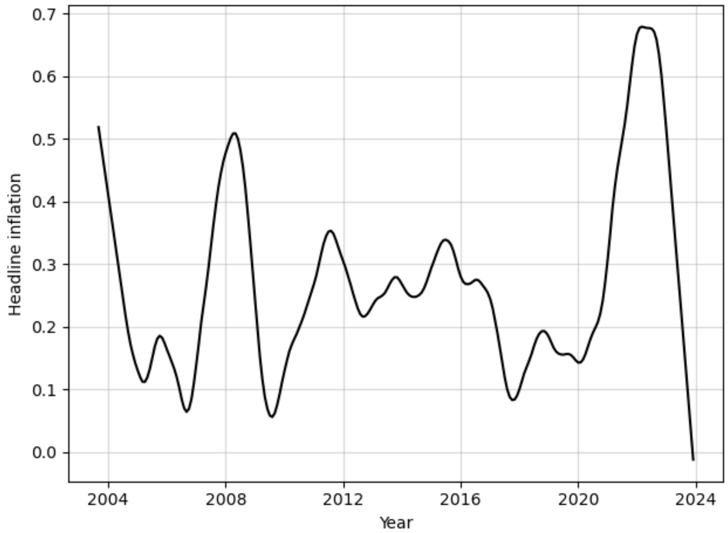
[55] Vega, M., Bigio, S., Florián, D., Llosa, G., Miller, S., Ramírez, N., Rodríguez, D., Salas, J., & Winkelried, D. (2009). Un Modelo Semi-estructural de Proyección para la Economía Peruana.*Revista Estudios Económicos, (19).* https://www.bcrp.gob.pe/docs/Publicaciones/Revista-Estudios-Economicos/17/Estudios-Economicos-17-2.pdf

[56] Winkelried, D. (2013). Modelo de Proyección Trimestral del BCRP: Actualización y novedades.*Revista Estudios Económicos, (26)*, 9–60.

[57] Wright, J. H. (2009). Forecasting US inflation by bayesian model averaging.*Journal of Forecasting, 28(2)*, 131–144. https://doi.org/10.1002/for.1088.

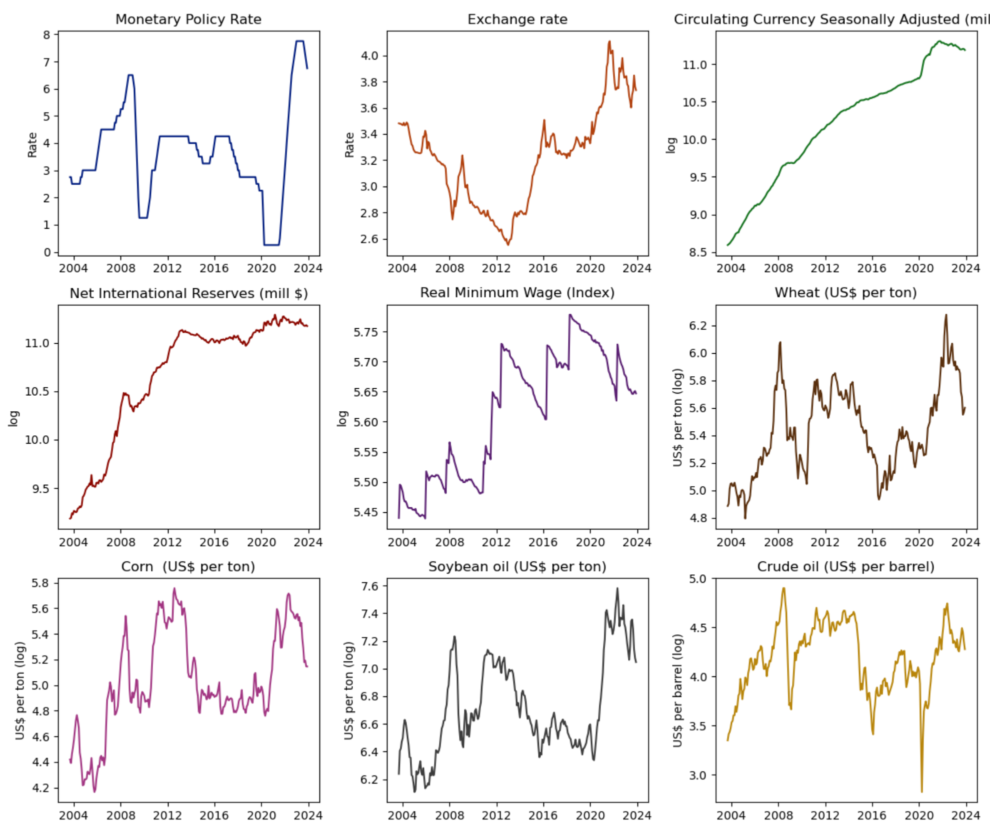
A Appendix

Figure A1: Headline Inflation



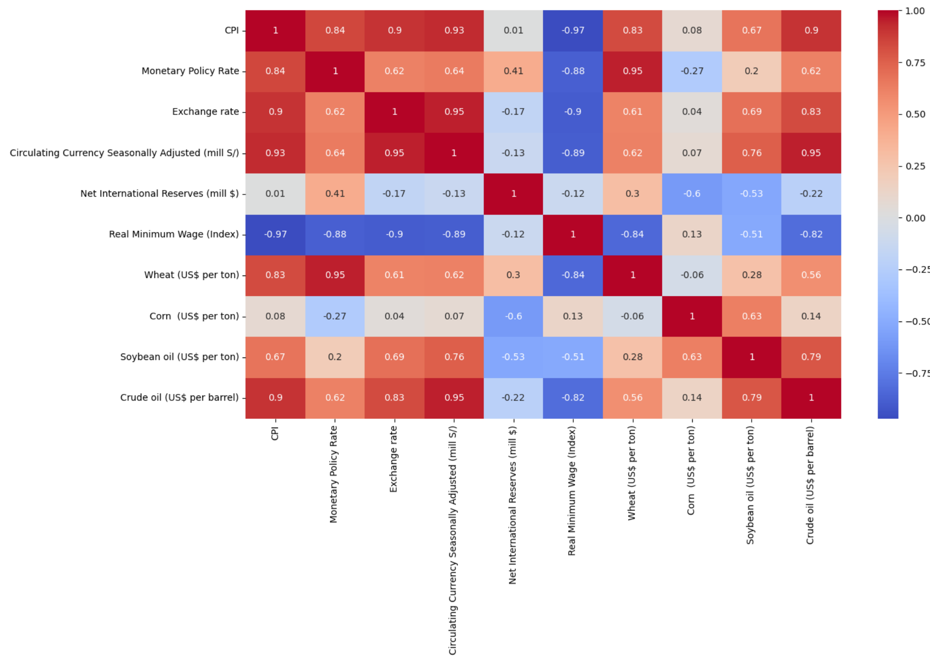
Headline Inflation (2003M9-2023M12)

Figure A2: Predictors for Headline Inflation



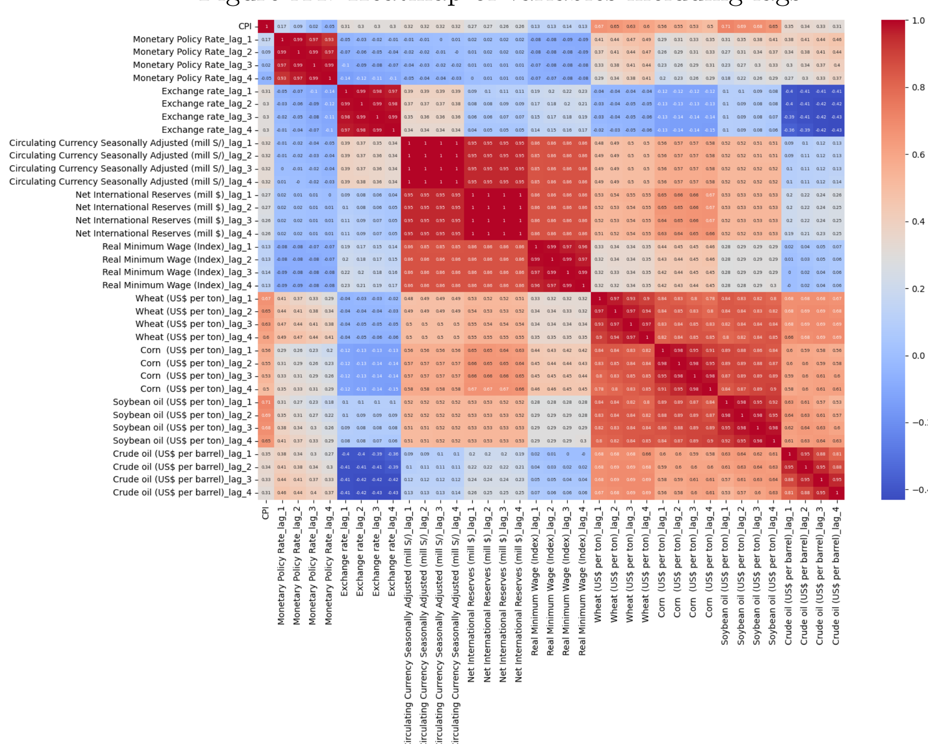
Variables used for the prediction of Headline Inflation (2003M9-2023M12)

Figure A3: Heatmap of variables



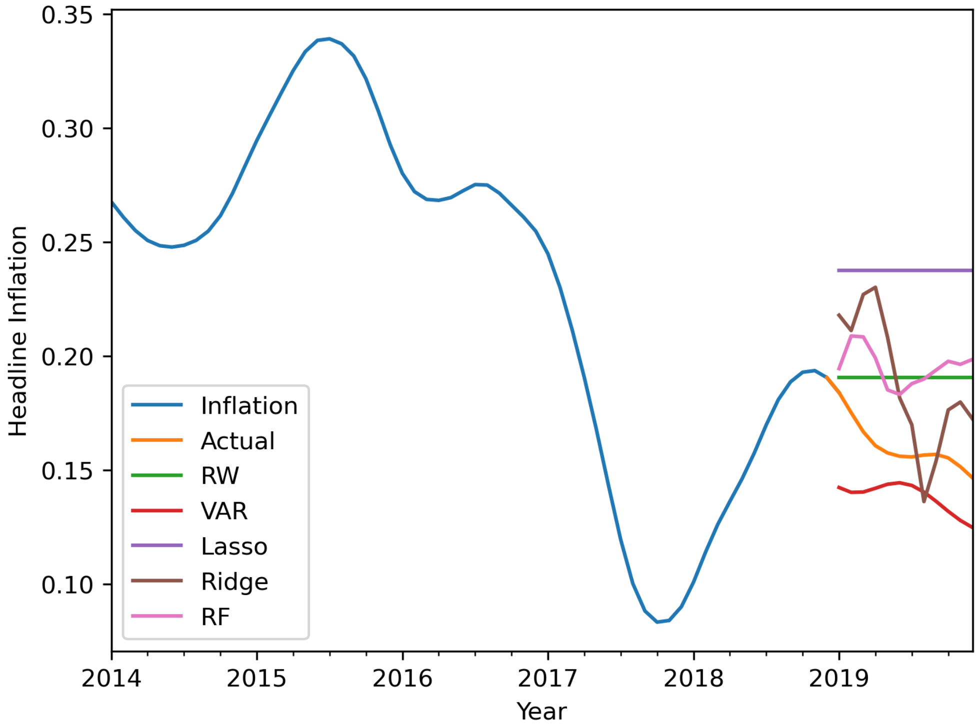
Heatmap correlation of headline inflation with all variables

Figure A4: Heatmap of variables including lags



Heatmap correlation of headline inflation all variables including lags

Figure A5: Comparison of forecasts for 2019



The orange line represents the actual inflation from 2019M1 to 2019M12.

Table A1: RMSE of forecasts (2019)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.007	0.012	0.017	0.021	0.024	0.026	0.027	0.028	0.029	0.030	0.031	0.032
VAR	6.378	**3.258**	2.083	1.519	1.218	**1.038**	**0.926**	**0.863**	**0.830**	**0.808**	**0.781**	**0.744**
Ridge	6.022	3.375	2.914	2.664	2.312	1.989	1.761	1.608	1.482	1.396	1.324	1.243
Lasso	8.233	4.945	3.750	3.204	2.929	2.775	2.684	2.634	2.600	2.565	2.516	2.451
RF	**2.008**	4.945	**1.598**	**1.350**	**1.182**	1.086	1.058	1.048	1.054	1.099	1.113	1.101

RMSE for Headline Inflation forecasts. The RMSE of other models is divided by the RMSE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A2: MAPE of forecasts (2019)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.035	0.062	0.088	0.113	0.132	0.147	0.158	0.165	0.171	0.176	0.184	0.193
VAR	6.378	3.458	2.200	1.552	**1.191**	**0.977**	**0.852**	**0.791**	**0.766**	**0.753**	**0.734**	**0.703**
Ridge	6.022	3.604	3.096	2.812	2.427	2.021	1.717	1.517	1.305	1.225	1.170	1.104
Lasso	8.233	5.274	4.049	3.445	3.123	2.935	2.818	2.748	2.700	2.654	2.597	2.530
RF	**2.008**	**1.981**	**1.655**	**1.412**	1.233	1.128	1.094	1.079	1.080	1.113	1.123	1.112

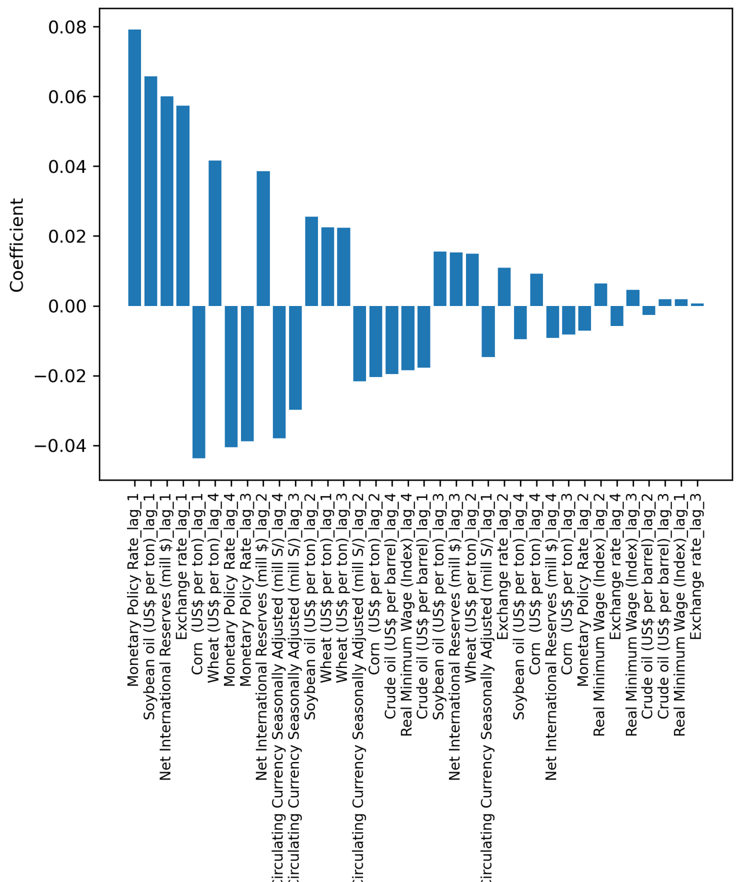
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A3: Diebold-Mariano test for 2019 forecast

Model	DM	p-value
VAR	0.864	0.406
Ridge	-0.479	0.641
Lasso	-9.083	0.000
RF	-1.068	0.308

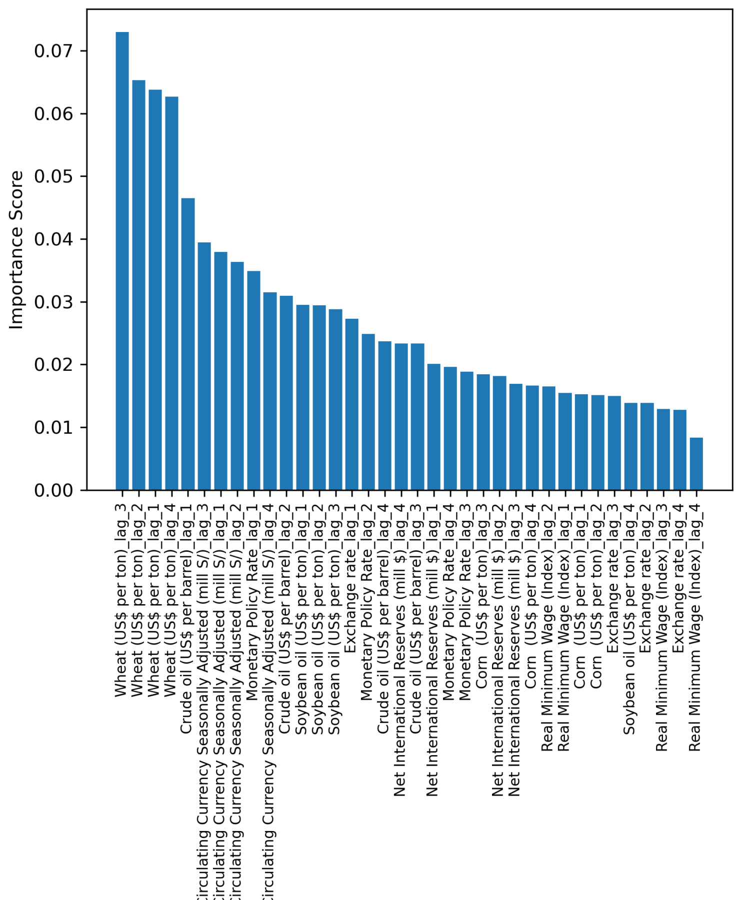
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A6: Main predictors of ridge regression



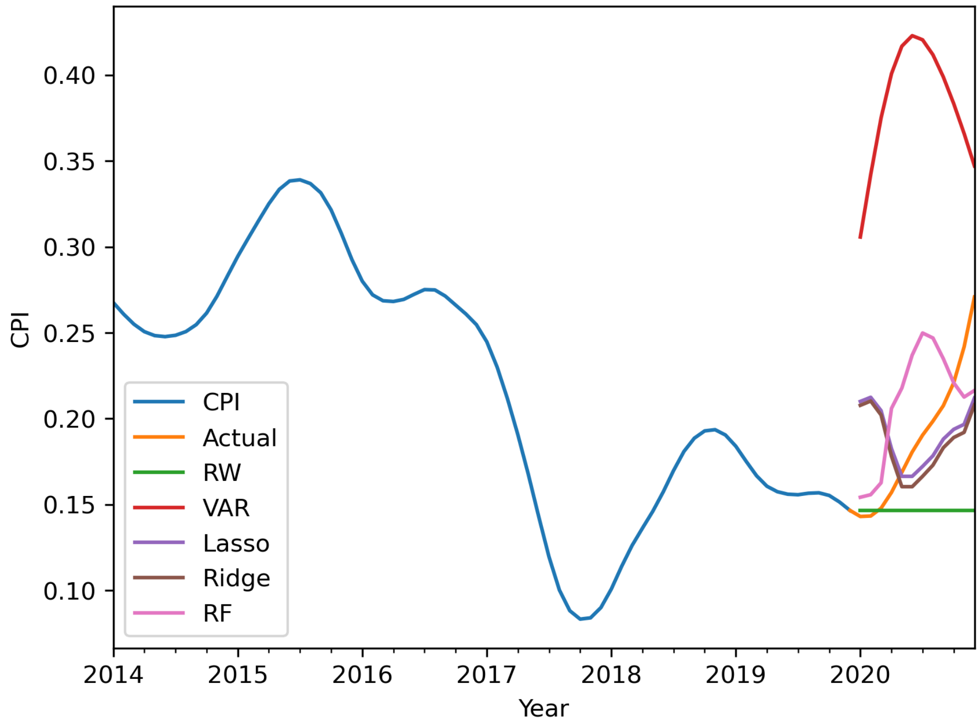
Main predictors chosen by the Ridge regression model for the 2019 forecast.

Figure A7: Main predictors of random forest regression



Main predictors chosen by the Random Forest regression model for the 2019 forecast.

Figure A8: Comparison of forecasts for 2020



The orange line represents the actual inflation from 2020M1 to 2020M12.

Table A4: RMSE of forecasts (2020)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.004	0.003	0.003	0.006	0.011	0.017	0.023	0.028	0.034	0.040	0.047	0.058
VAR	46.077	53.393	68.826	36.359	19.535	12.910	9.722	7.858	6.533	5.414	4.381	3.458
Ridge	18.327	19.355	21.627	9.489	4.404	2.651	1.881	**1.466**	**1.191**	0.991	0.850	0.737
Lasso	18.968	20.007	22.427	9.905	4.585	2.735	1.921	1.482	1.195	0.985	0.835	0.718
RF	**3.169**	**3.477**	**4.480**	**4.647**	**2.911**	**2.180**	**1.801**	1.497	1.221	**0.982**	**0.804**	**0.686**

RMSE for Headline Inflation forecasts. The RMSE of other models is divided by the RMSE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A5: MAPE of forecasts (2020)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.025	0.024	0.019	0.031	0.051	0.074	0.096	0.117	0.136	0.156	0.178	0.201
VAR	46.077	53.162	72.180	45.720	27.874	19.036	14.319	11.456	9.466	7.902	6.576	5.445
Ridge	18.327	19.366	22.864	11.572	5.785	3.575	2.539	1.966	1.591	1.342	1.177	1.049
Lasso	18.968	20.018	23.720	12.193	5.945	3.591	2.504	1.966	1.531	1.280	1.118	0.995
RF	**3.169**	**3.474**	**4.709**	**4.696**	**3.412**	**2.665**	**2.217**	**1.857**	**1.520**	**1.193**	**1.015**	**0.905**

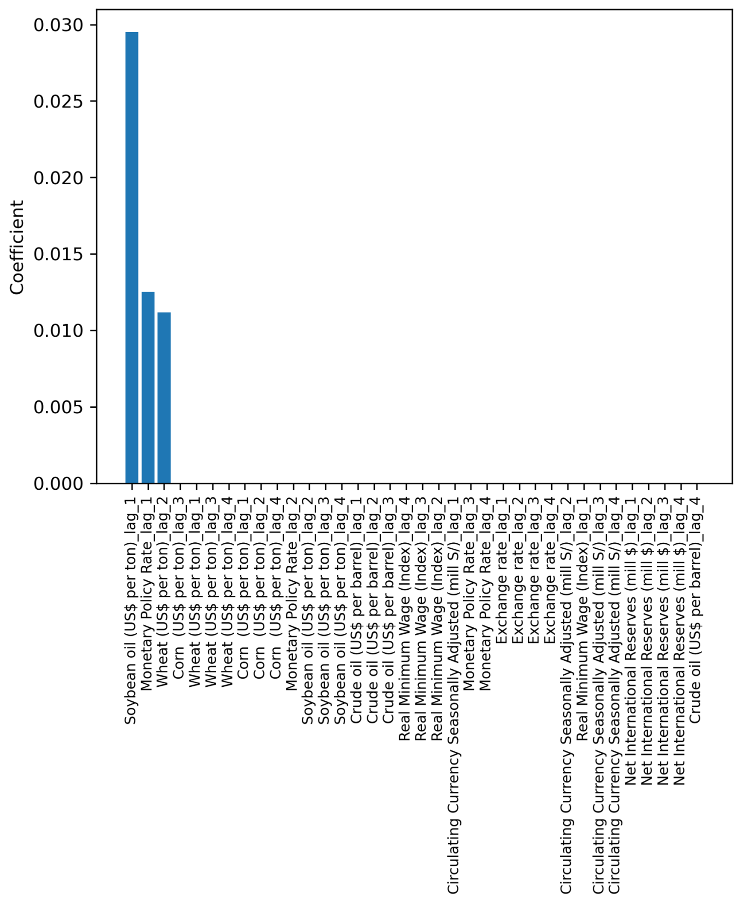
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A6: Diebold-Mariano test for 2020 forecast

Model	DM	p-value
VAR	-2.689	0.021
Ridge	0.556	0.589
Lasso	0.553	0.591
RF	0.618	0.549

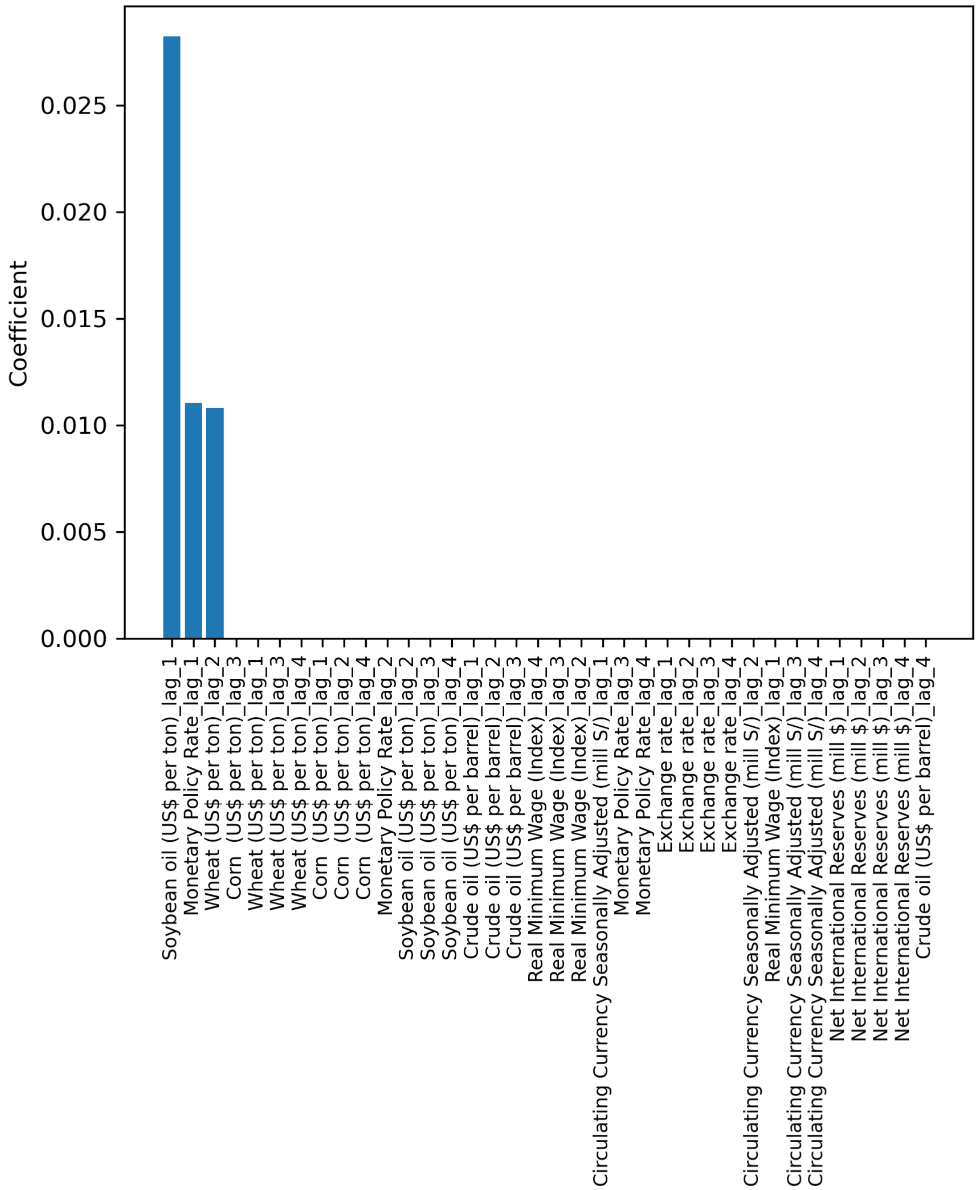
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A9: Main predictors of ridge regression



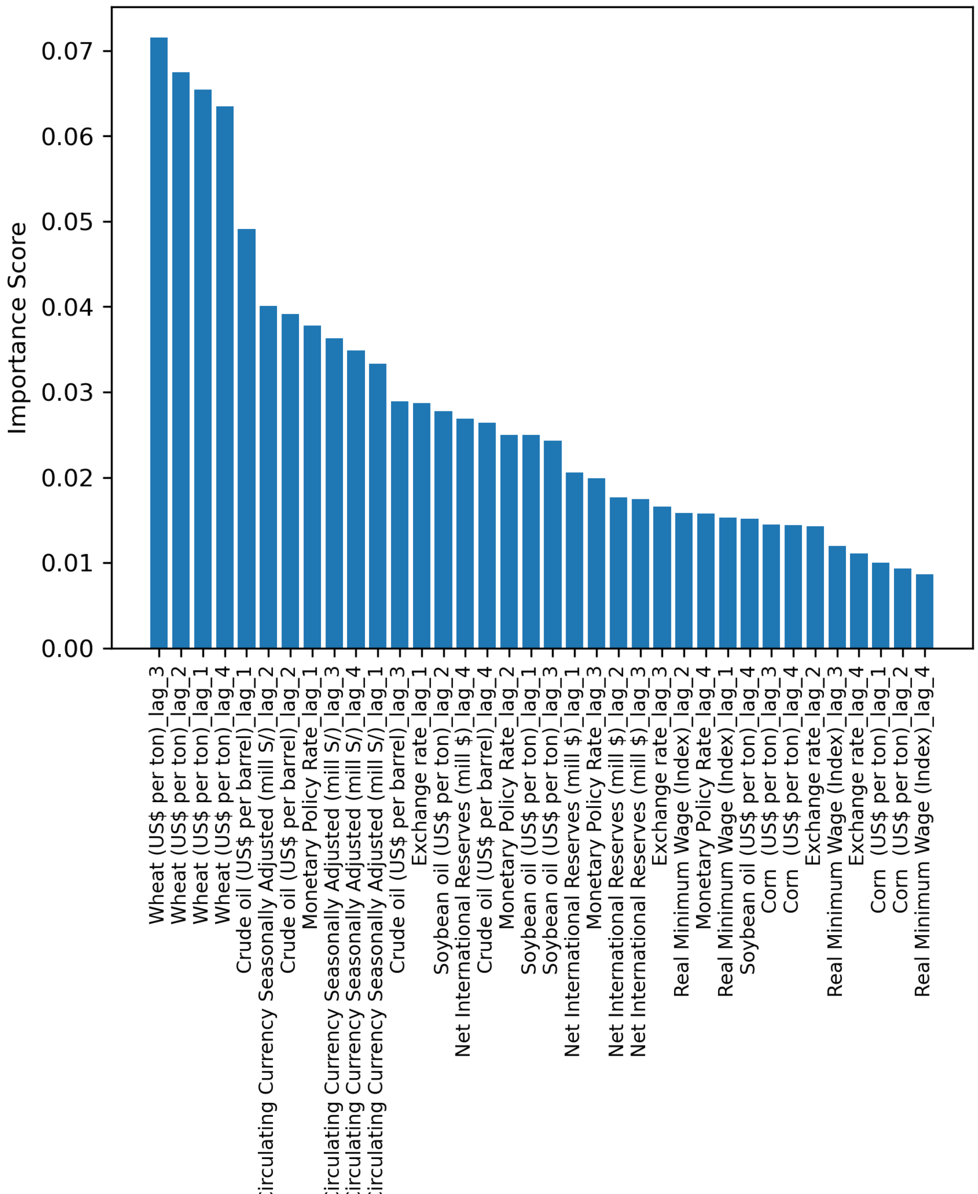
Main predictors chosen by the ridge regression model for the 2020 forecast.

Figure A10: Main predictors of LASSO regression



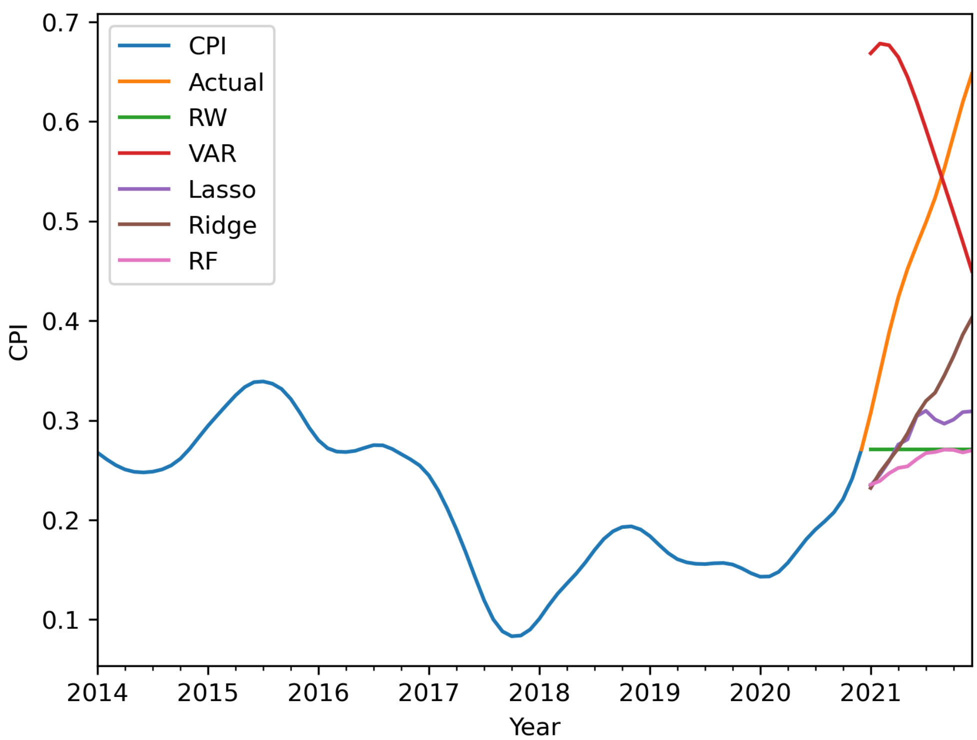
Main predictors chosen by the LASSO regression model for the 2020 forecast.

Figure A11: Main predictors of random forest regression



Main predictors chosen by the random forest regression model for the 2020 forecast.

Figure A12: Comparison of forecasts for 2021



The orange line represents the actual inflation from 2021M1 to 2021M12.

Table A7: RMSE of forecasts (2021)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.036	0.060	0.084	0.105	0.124	0.141	0.156	0.171	0.187	0.203	0.221	0.238
VAR	9.929	5.760	3.925	2.934	2.330	1.920	1.619	1.384	1.196	1.050	0.943	0.872
Ridge	2.063	**1.471**	**1.237**	1.115	**1.033**	**0.967**	**0.916**	**0.880**	**0.846**	**0.814**	**0.784**	**0.757**
Lasso	1.993	1.474	1.242	**1.106**	1.040	0.974	0.932	0.919	0.916	0.913	0.909	0.907
RF	**1.971**	1.529	1.325	1.222	1.169	1.128	1.095	1.073	1.055	1.042	1.035	1.028

RMSE for Headline Inflation forecasts. The RMSE of other models is divided by the RMSE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A8: MAPE of forecasts (2021)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.118	0.170	0.214	0.251	0.281	0.306	0.327	0.347	0.365	0.382	0.399	0.414
VAR	9.929	6.266	4.475	3.433	2.756	2.274	1.903	1.600	1.360	1.204	1.101	1.033
Ridge	2.063	1.567	**1.344**	1.217	**1.130**	**1.060**	**1.005**	**0.965**	**0.930**	**0.898**	**0.869**	**0.843**
Lasso	1.993	**1.561**	1.345	**1.209**	1.133	1.064	1.017	0.993	0.980	0.970	0.960	0.953
RF	**1.971**	1.607	1.417	1.311	1.249	1.202	1.165	1.138	1.116	1.100	1.088	1.078

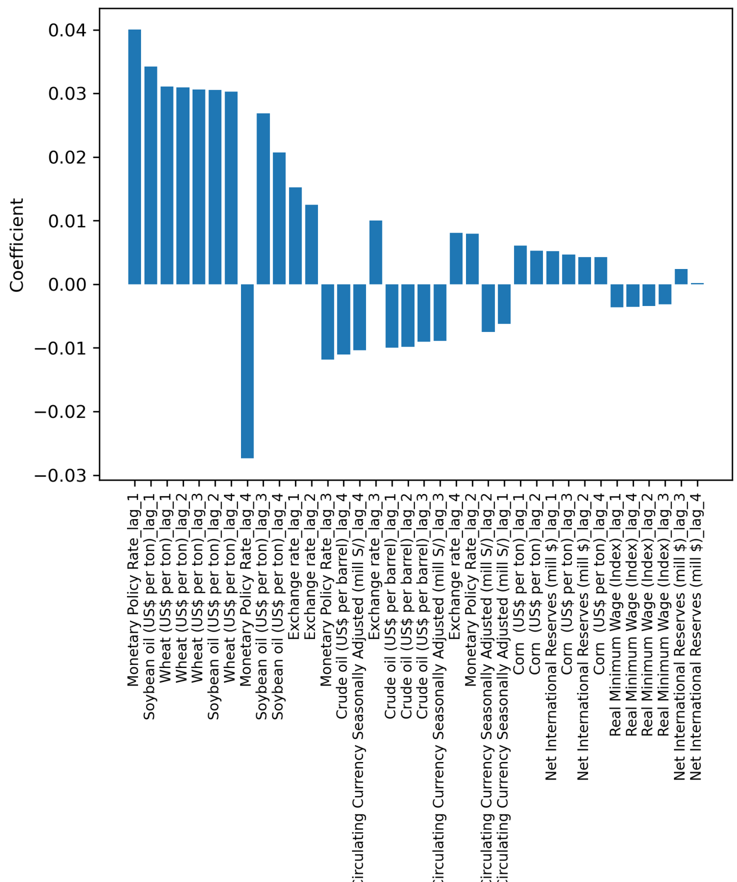
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A9: Diebold-Mariano test for 2021 forecast

Model	DM	p-value
VAR	0.252	0.806
Ridge	1.284	0.226
Lasso	1.417	0.184
RF	-1.956	0.076

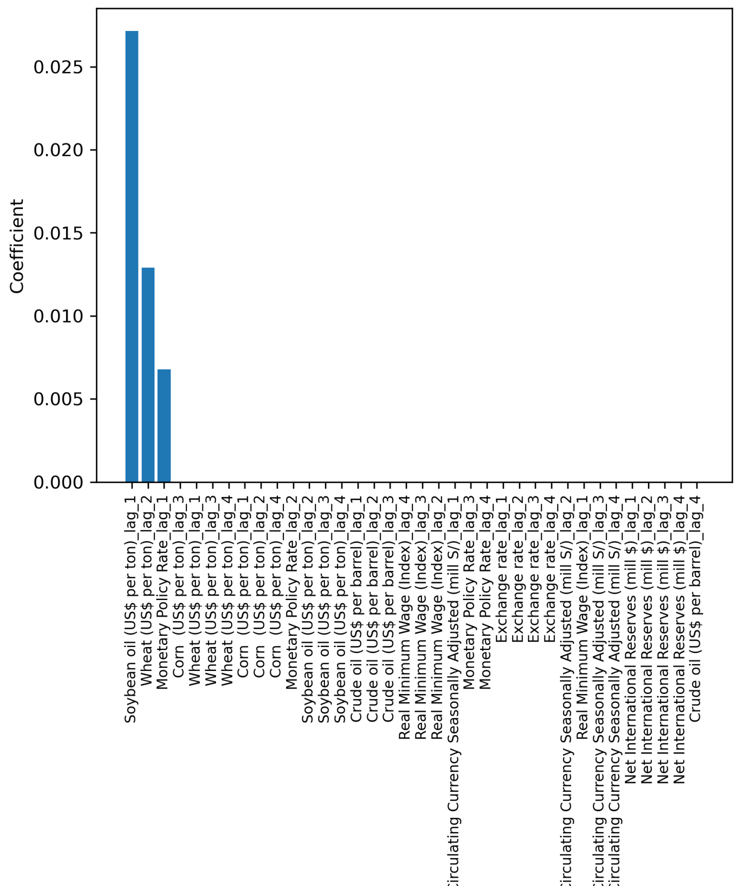
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A13: Main predictors of ridge regression



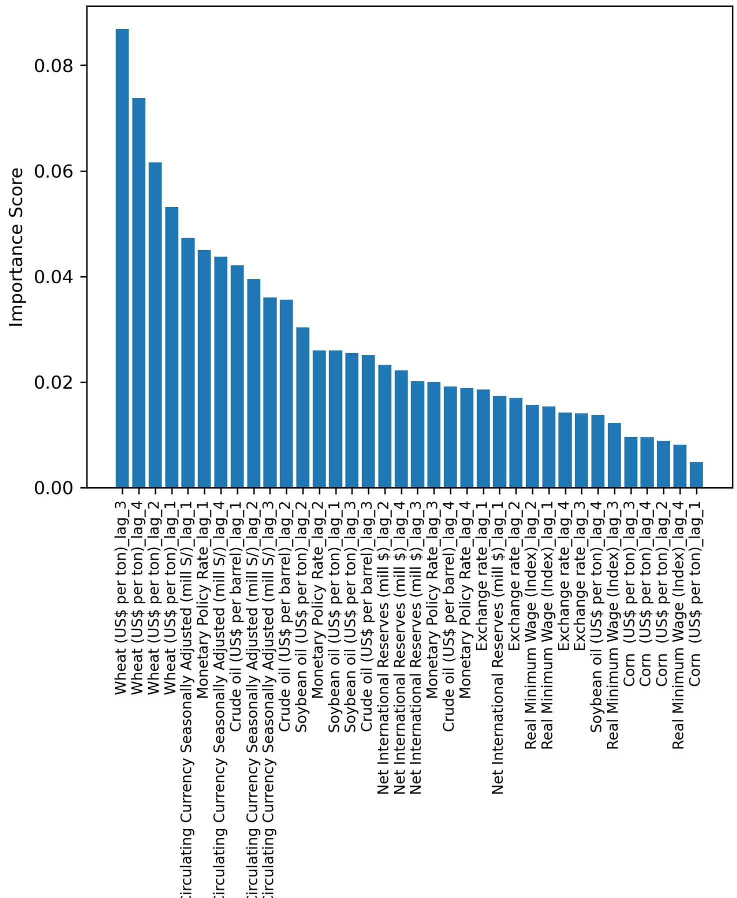
Main predictors chosen by the ridge regression model for the 2021 forecast.

Figure A14: Main predictors of LASSO regression



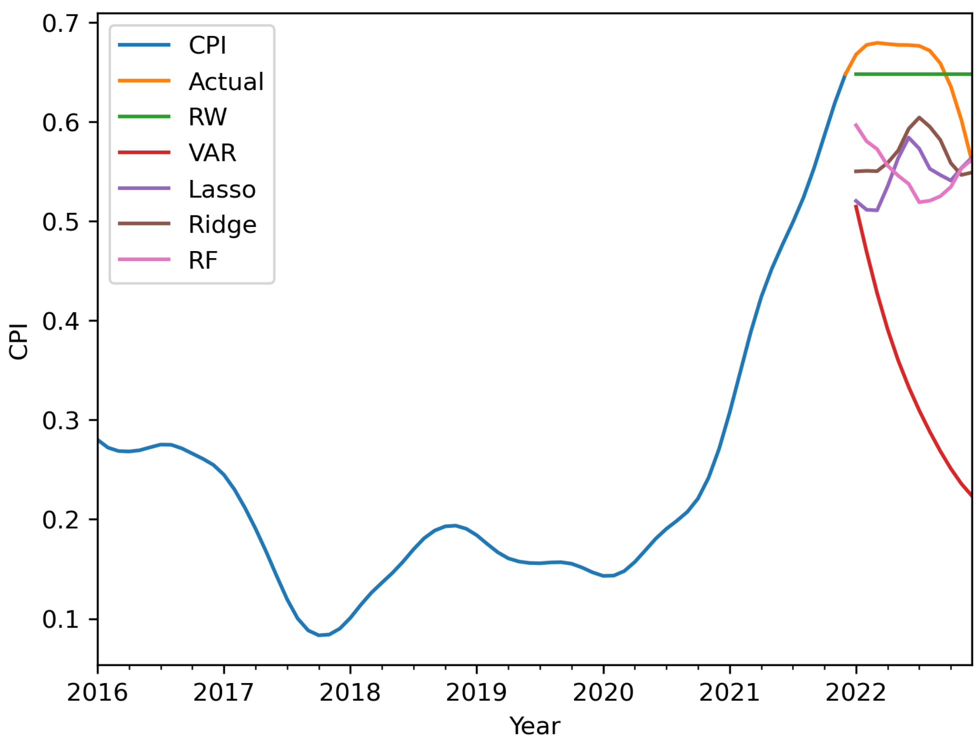
Main predictors chosen by the LASSO regression model for the 2021 forecast.

Figure A15: Main predictors of random forest regression



Main predictors chosen by the random forest regression model for the 2021 forecast.

Figure A16: Comparison of forecasts for 2022



The orange line represents the actual inflation from 2022M1 to 2022M12.

Table A10: RMSE of forecasts (2022)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.020	0.025	0.027	0.028	0.028	0.028	0.028	0.028	0.026	0.025	0.028	0.037
VAR	7.793	7.321	7.644	8.227	8.851	9.436	10.010	10.721	11.701	12.505	11.543	8.792
Ridge	5.989	4.901	4.573	4.403	4.248	4.044	**3.866**	**3.815**	**3.902**	**3.971**	**3.496**	**2.542**
Lasso	7.504	6.288	5.905	5.590	5.270	4.970	4.805	4.827	4.987	5.064	4.424	3.214
RF	**3.624**	**3.409**	**3.409**	**3.608**	**3.814**	**4.003**	4.258	4.493	4.759	4.865	4.253	3.090

RMSE for Headline Inflation forecasts. The RMSE of other models is shown as the ratio of the RMSE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A11: MAPE of forecasts (2022)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.029	0.036	0.040	0.041	0.041	0.042	0.042	0.041	0.038	0.036	0.040	0.050
VAR	7.793	7.380	7.634	8.143	8.705	9.240	9.773	10.472	11.700	12.727	11.890	9.775
Ridge	5.989	4.998	4.658	4.467	4.291	4.049	**3.833**	**3.771**	**3.932**	**4.051**	**3.554**	**2.654**
Lasso	7.504	6.405	6.007	5.660	5.289	4.928	4.744	4.777	5.047	5.181	4.460	3.299
RF	**3.624**	**3.435**	**3.426**	**3.597**	**3.783**	**3.957**	4.187	4.426	4.807	4.982	4.297	3.174

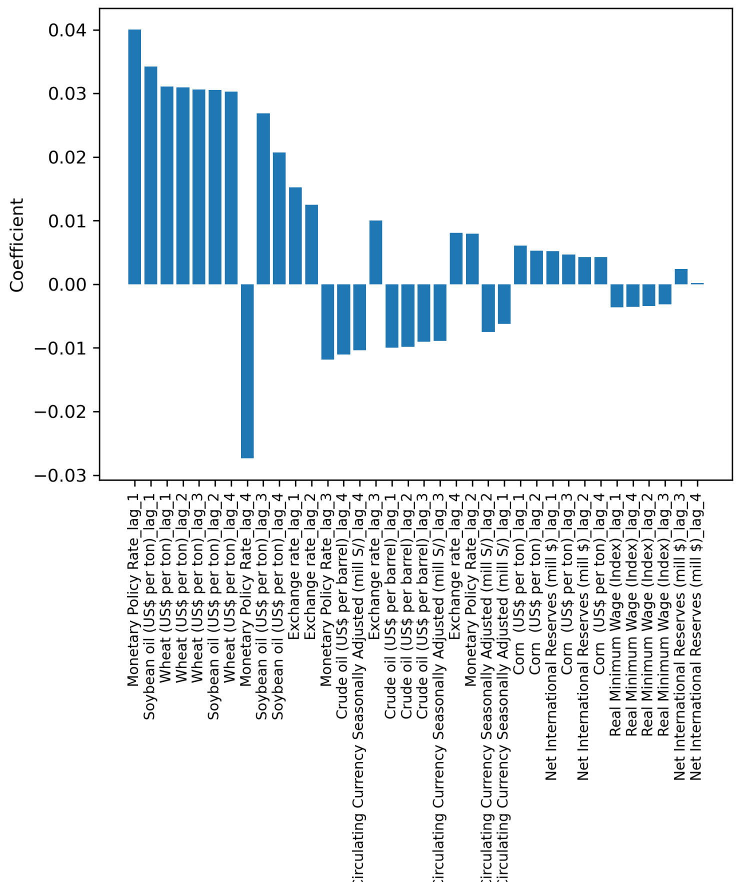
MAPE for Headline Inflation forecasts. The MAPE of other models is shown as the ratio of the MAPE of the benchmark RW model. The best model for each horizon is outlined in bold.

Table A12: Diebold-Mariano test for 2022 forecast

Model	DM	p-value
VAR	-3.684	0.004
Ridge	-1.884	0.086
Lasso	-2.144	0.055
RF	-2.233	0.047

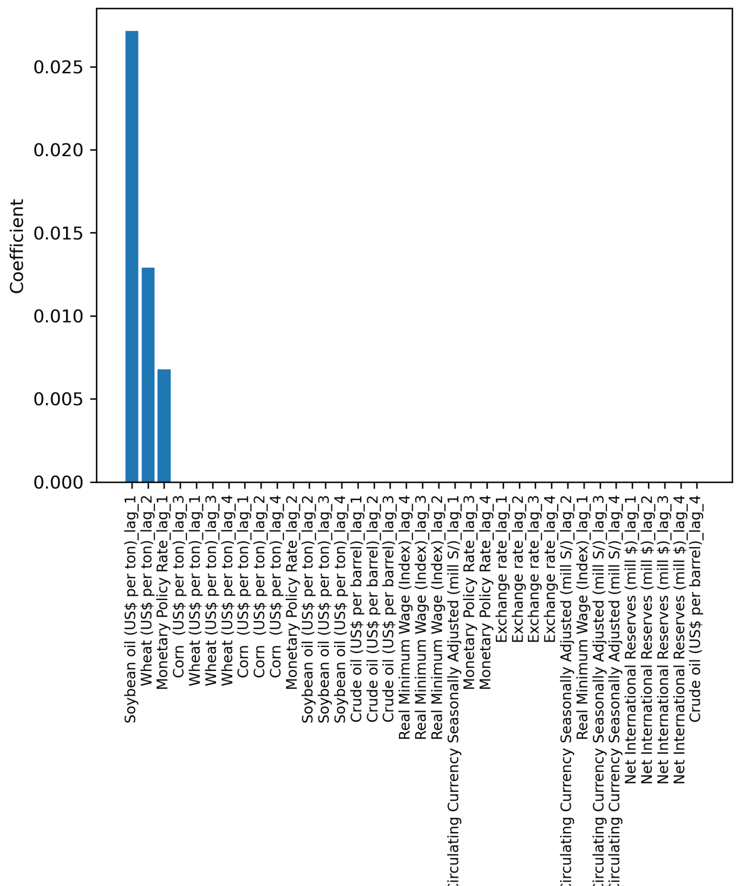
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A17: Main predictors of ridge regression



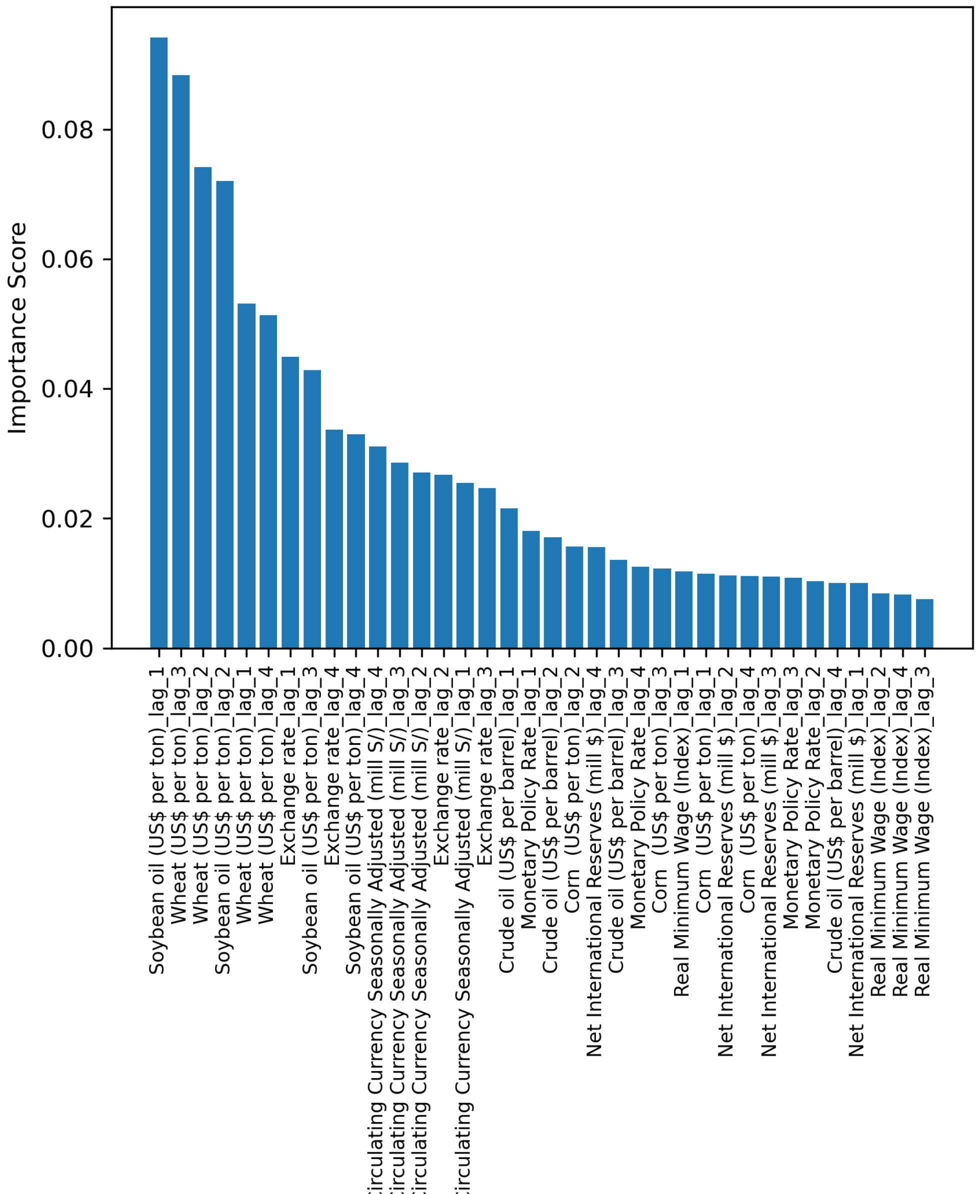
Main predictors chosen by the ridge regression model for the 2022 forecast.

Figure A18: Main predictors of LASSO regression



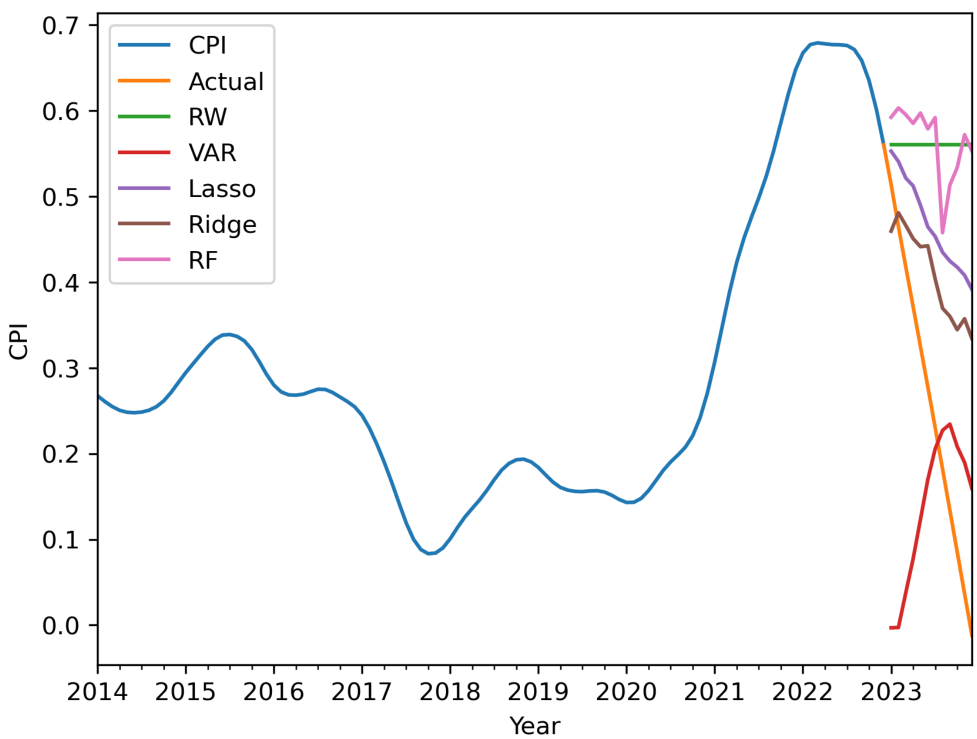
Main predictors chosen by the LASSO regression model for the 2022 forecast.

Figure A19: Main predictors of random forest regression



Main predictors chosen by the random forest regression model for the 2022 forecast.

Figure A20: Comparison of forecasts for 2023



The orange line represents the actual inflation from 2023M1 to 2023M12.

Table A13: RMSE of forecasts (2023)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.046	0.075	0.102	0.129	0.156	0.183	0.211	0.238	0.266	0.293	0.321	0.349
VAR	12.505	7.631	5.529	4.319	3.514	2.930	2.482	2.127	1.838	1.600	1.403	1.238
Ridge	1.773	**1.398**	**1.177**	**1.017**	**0.922**	**0.865**	**0.832**	**0.822**	**0.821**	**0.819**	**0.810**	**0.798**
Lasso	**1.728**	1.415	1.221	1.081	1.000	0.933	0.909	0.915	0.929	0.936	0.932	0.918
RF	2.098	1.707	1.503	1.379	1.225	1.095	1.005	0.958	0.934	0.923	0.928	0.920

RMSE for Headline Inflation forecasts. The RMSE of other models is divided by the RMSE of the benchmark RW model. The best model is outlined in bold.

Table A14: MAPE of forecasts (2023)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark	0.089	0.147	0.211	0.286	0.373	0.480	0.616	0.798	1.063	1.513	2.671	6.473
VAR	12.505	7.942	5.780	4.478	3.583	2.910	2.372	1.920	1.522	1.149	**0.748**	**0.383**
Ridge	1.773	**1.441**	**1.223**	**1.058**	**0.954**	**0.889**	**0.849**	**0.832**	**0.828**	**0.822**	0.803	0.772
Lasso	**1.728**	1.452	1.262	1.119	1.032	0.957	0.925	0.926	0.939	0.945	0.933	0.891
RF	2.098	1.753	1.551	1.421	1.261	1.119	1.017	0.961	0.932	0.916	0.930	0.906

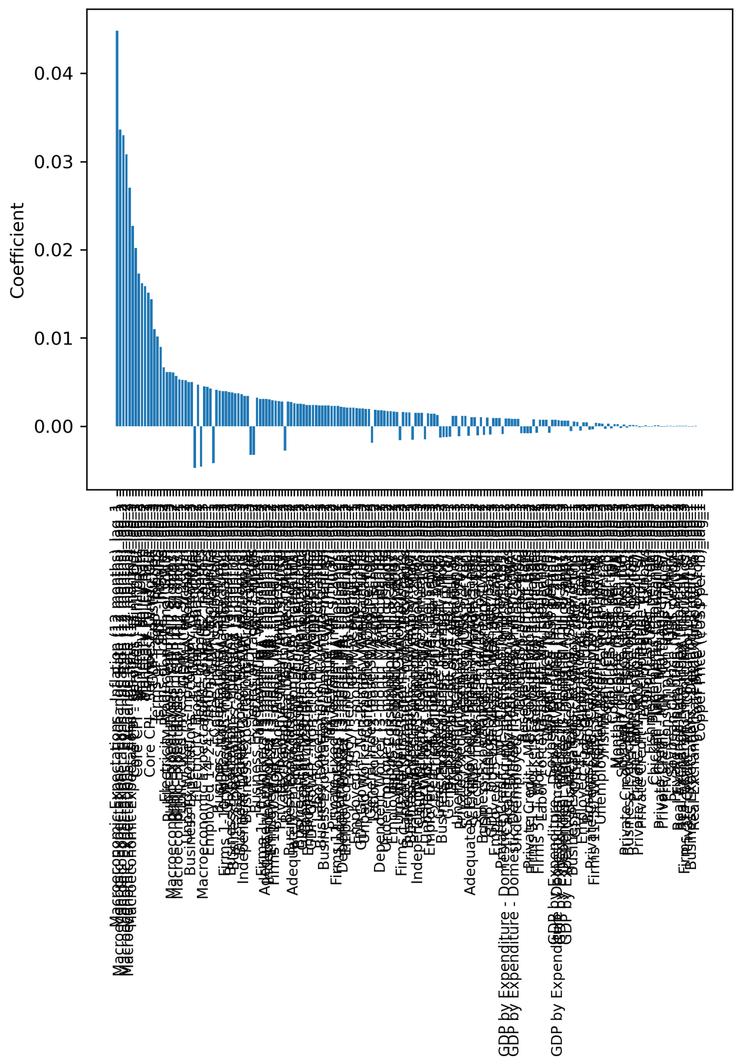
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model is outlined in bold.

Table A15: Diebold-Mariano test for 2023 forecast

Model	DM	p-value
VAR	-0.431	0.675
Ridge	1.533	0.154
Lasso	1.510	0.159
RF	0.965	0.355

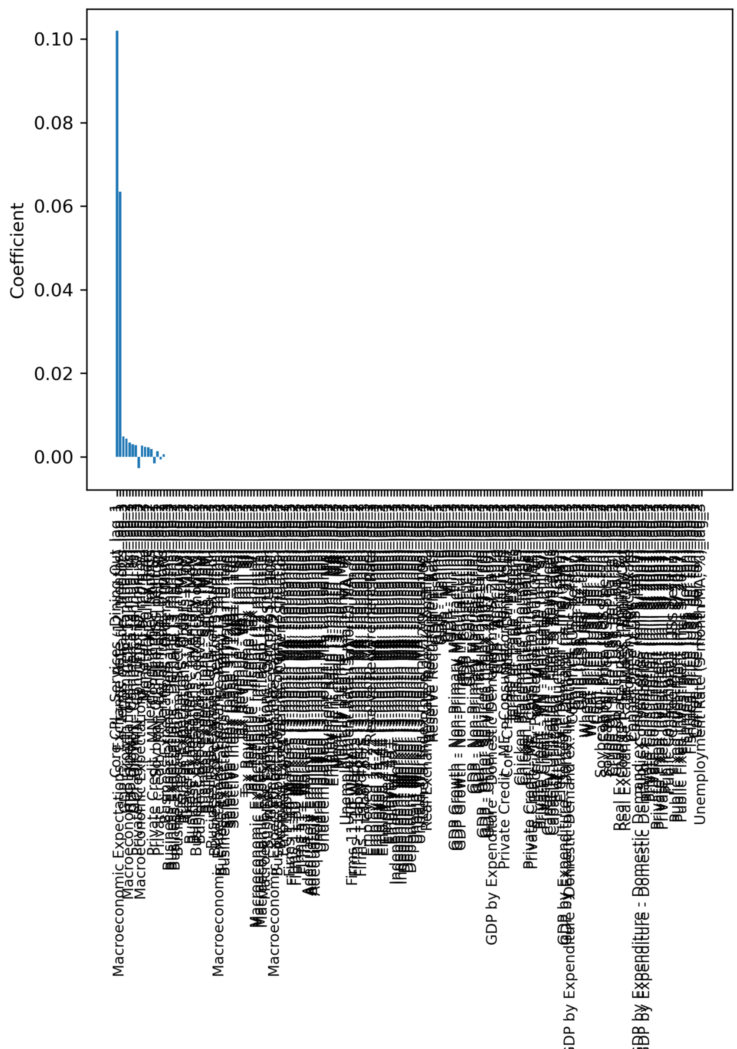
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A21: Main predictors of ridge regression



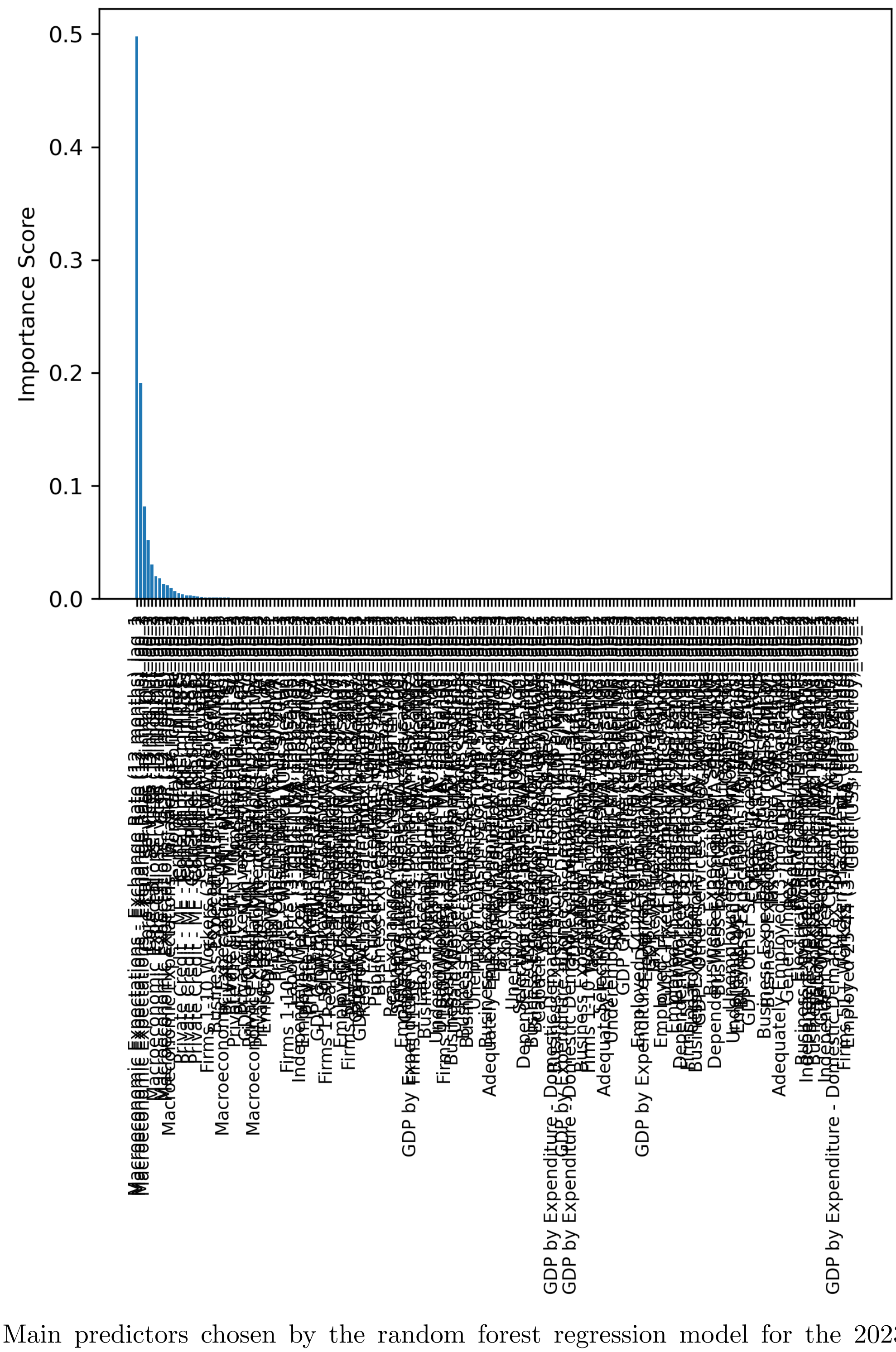
Main predictors chosen by the ridge regression model for the 2023 forecast.

Figure A22: Main predictors of LASSO regression



Main predictors chosen by the LASSO regression model for the 2023 forecast.

Figure A23: Main predictors of random forest regression



Main predictors chosen by the random forest regression model for the 2023 forecast.