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: April 8, 2024

Abstract

Inflation forecasting is a key element of monetary policy. Bayesian techniques have been used to improve upon univariate methods. 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. Our goal is to demonstrate that both linear and non-linear machine learning models can easily outperform traditional models, and even Bayesian techniques in a in an emerging economy with a steady price level like Peru. 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.

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 (Ülke et al., 2018; Medeiros et al., 2022). In particular, the find out that non-linear models, like Random Forest, are way better at predicting inflation than 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.

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 recently done in Peru (Flores & Grandez, 2023).

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 Autoregressive Integrated Moving Average (ARIMA) and Vector Autoregression (VAR), while the ML methods will be a LASSO regression, Ridge regression, Elastic Net (EN) and Random Forest (RF). We will see which models perform better during three main forecasting periods: January 2009 - December 2009, January 2019 - December 2019 and January 2023 - December 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.

The structure of the document includes the following: Section 2 outlines both the theoretical and empirical literature regarding inflation forecasting; Section 3 describes the theoretical framework that is going to be used to compare the forecasting models; Section 4 portrays the data and describes the models; Section 5 comprises the discussion of the main results, followed by the concluding section.[1]

2 Literature review

In this section we will review the main literature regarding inflation and time-series forecasting. Subsection 2.1 contains a brief describe of the evolution of forecasting theory and forecasting evaluation criteria. Subsection 2.2 examines different forecasting methods that has been used in empirical works, from simple univariate ARMA models, to State-Space representations, Stochastic Volatility models and Bayesian Vector Autoregressions. It first shows international documentation, then regional literature followed by papers centred in Peruvian inflation. It then reviews the most recent method: Machine Learning models, and how they have been used to forecast inflation in countries in the region.

2.1 Theorical Literature

The theory of economic forecasting has roots in the beginning of the twentieth century, with the first extensive work being that of Morgenstern (1928). According to Clements and Hendry (1998), based on reviews by Marget (1929), his work was the first one to discuss in-depth economic and business forecasting. He argued against it since economic data was not suitable for forecasting (it was not homogenous, nor independently distributed, and samples were too small to be used) and forecasting itself could impact agents' reactions to them, making it ineffective (this was some sort of first 'Lucas's critique').

Marget, on the other hand, while agreeing with the first argument of Morgenstern, stated that forecasting should be viable as long as it was based on the extrapolation of previous patterns, rather than on probability-based techniques. This idea would imply that causation means predictability, which is not always the case. Marget would also support the idea of testing economic theories using forecasting. Tinbergen (1939) was one of the first to do so, by constructing forecasting tests for econometric models developed for two reports delegated by the League of Nations. The role of

[1] These are preliminary conclusions, subject to change.

econometric modelling and forecasting was later enhanced by the work of Theil (1961), Heesterman (1970), and Pindyck & Rubinfeld (1981), among others.

Another endeavour worth acknowledging is Haavelmo (1944). Haavelmo had a probabilistic approach to econometrics, proclaiming that the nature of economic data shouldn't be an obstacle to applying statistical tools to its analysis. His forecasting approach (with a probabilistic framework) was the precursor of the textbook economic forecasting (Clements & Hendry, 1998). It is worth noting that his analysis required certain regularity in the series being forecasted (e.g., no structural breaks) since alternative forecasting techniques to deal with breaks in the data were yet to be developed.

The forecast evaluation criteria changed and developed during this period. The work of Klein (1950), Brown (1954) and Theil (1961) had a predominant role in the development of forecasting theory in the post-war period. Brown derived standard errors for forecasts from systems, while Theil developed the Theil index, a comparison measure for forecasting. Granger and Newbold (1973) showed that forecast accuracy cannot be based on comparisons of the predicted and actual time series or their distributional properties. Instead, we should focus on their difference, the forecast error. Ericsson (1992) considers the traditional use of the mean square forecast error (MSFE) in forecast evaluation, as well as the consistency of the models being compared.

2.2 Empirical Literature

Based on the framework developed by Box and Jenkins (1976), most literature on forecasting in the post-war period focused on ARMA and ARIMA models (Anderson, 1977; Harvey & Todd, 1983). The Wold decomposition theorem (Wold, 1938) postulates that a purely deterministic covariance stationary process can have an infinite moving-average (MA) representation. Furthermore, any MA(∞) can be represented in accuracy by an ARMA(p,q) model. Due to the stochastic nature of most economic time series, they require differencing at least once in order to be stationary. For this reason, the integrated autoregressive moving average (ARIMA(p,d,q)) models, in which the d indicates the number of times the series has been differentiated to be stationary, became dominant in modelling and forecasting.

Time series have also taken the form proposed by Kalman (1960), who designed a technique to identify and extract latent factors in state-space models. These models are useful since they relax the assumption that the explanatory variable is observable, which can help to explain the behaviour of some economic and financial time series. This allows for an easy way to model dynamic multivariate time series and avoids the need for proxy variables (Martin et al., 2013). The unobserved components approach has been proven to be a good fit for modelling and forecasting, competing and outperforming ARIMA models (Harvey & Todd, 1983; Stock & Watson, 1988, 1989, 1993, 2003). However, it has also been pointed out that when using UC models, there can be issues both in the specification of the models and in the inferences drawn from the distribution of the unobserved components (Maravelli, 1994).

A more recent approach of the unobserved components model with stochastic volatility (UC-SV) has been used to forecast inflation (Stock & Watson, 2007). Considering that inflation has both a permanent stochastic trend component that has suffered large changes in its variance since the post-war period and a serially uncorrelated transitory component with a relatively constant variance, Stock & Watson compare different models, including a univariate AR model, an Atkeson–Ohanian (2001) random walk, an integrated moving-average IMA(1,1), multiple multivariate Backwardlooking Phillips curve (PC) and a UC-SV model. They divide the US inflation data into two samples (1970Q1- 1983Q4 and 1984Q1-2004Q4) to produce pseudo-out-of-sample forecasts. They concluded that in the second period, MSFE has declined substantially, meaning that inflation has become relatively easier to forecast. However, the traditional AR and PC models have reduced their performance, whereas the UC-SV and IMA models appear to hold the upper hand in forecasting.

Another approach to forecasting was developed in the 80s, embracing Bayesian methods for estimating vector autoregressions (Litterman, 1980; Litterman, 1986; Doan et al., 1984; Sims, 1993) showing that the Bayesian vector autoregression (BVAR) approach had better results than their univariate counterparts. Litterman was a pioneer in forecasting macroeconomic variables based on Bayesian shrinkage estimation techniques developed in previous decades (Shiller, 1970; Stein, 1960). Doan, Litterman, and Sims provided unconditional forecasts 1982:12-1983:3 for the U.S. Congressional Budget Office and described how these Bayesian techniques with time-varying parameters could be useful for conditional projections and to discuss policy alternatives. Sims modified Litterman's model to allow for conditional heteroskedasticity and non-normality of disturbances and used it to forecast nine macroeconomic variables. His model worked pretty well with unconditional predictions, but it felt short with conditional predictions, therefore, was not as useful to analyze policy interventions.

The adoption of the Bayesian techniques would lead to an increase in literature that would use TVP in modelling and forecasting (Canova, 1993; Cogley and Sargent, 2001; Cogley and Sargent, 2002; Primiceri, 2005). Canova asserts that due to the short-term changes in financial time series, they are better described by stochastic processes that have leptokurtic distributions (with heavy tails and a sharper peak) and are conditionally heteroskedastic. In that sense, the author proposes as a forecasting tool a Bayesian timevarying coefficient (TVP-BVAR) model. HAssessing forecasting in linear and non-linear Machine Learning models for short-term inflation in Perue forecasted the U.S. spot exchange rates and spot interest rates from 1986.1 to 1987.51 and concluded that the TVP model improves upon RW forecasts at all horizons and for all the exchange rates. Similarly, Cogley and Sargent used a non-linear TVP-BVAR model to describe inflation-unemployment dynamics in the post-war U.S. economy. They then extended the model to include stochastic volatility to capture the drift in the persistence of inflation. Primiceri also used a TVP model to capture policy changes in a structural vector autoregression (TVP-SVAR) and found evidence of time variation in the U.S. monetary policy.

Most of the research and breakthroughs in inflation estimation and forecasting have been focused on developed countries. However, in the last few decades, literature concerning forecasting in the region has also implemented new techniques. One example is Gil-Alana & Pestana Barros (2009) who used a Fractional-Cointegrated VAR (FCVAR) with structural break to analyse inflation in 18 Latin American countries. They found out that most countries had a break date around the late 80s and early 90s. For the FCVAR model, they were inspired by Bos et. al (2001), who reviewed the forecasting ability of fractional models.

Similarly, the most recent work of Pincheira-Brown et al. (2019) examines the ability of core inflation to predict consumer price index (CPI) annual inflation in eight Latin American countries. They used a sample period (January 1995–May 2017) where most countries faced changes in their monetary regimes and different international and domestic financial crisis. Using two nested VAR models, they find out the core inflation is an important predictor for headline inflation in the short-run, but does not add much information for most countries in the long-run.

In Peru, there is also literature that has used these innovative approaches to forecast inflation.

Llosa et al. (2006) were pioneers in the introduction of Bayesian techniques following the methodology of Doan et al. (1984) to forecast Peruvian inflation. The authors used four specifications for BVARs (which included from five (the simplest) to nine variables) using 6 lags for all models. The evaluation of the predictive accuracy of each BVAR specification is obtained by estimating and recursively predicting. The estimation period ranges from August 1995 to November 2001 and the forecast period goes from December 2001 to August 2004. To evaluate the out-of-sample forecast of the different models, they used both the mean square error (MSE) and the U-Theil statistic.

In general, they find that all models except the BVAR 4 perform quite well in predicting inflation in the first out-of-sample period and, on average, the BVAR 1 has the best inflation forecasting performance with the lowest U-Theil statistic (0.16). When it comes to robustness, the BVAR 1 also outperforms all other models. Therefore, they conclude that out-of-sample predictions made with BVARs favour small BVAR specifications under different possible criteria.

Vega et al. (2009) and Wilkelried (2013) describe the quarterly macroeconometric model, known as Modelo de Proyección Trimestral (MPT), used by the BCRP as a reference to make policy simulations and projections of key macroeconomic variables in Peru. The MPT is a dynamic semi-structural model used to explain the aggregate behaviour of a small open economy with partial dollarisation. It uses linear behavioural equations obtained from non-linear equations derived from the optimal behaviour of economic agents subject to preferences, technologies, and market structures. The MPT has four main blocks (1) Aggregate supply, (2) Aggregate demand, (3) Uncovered interest rate parity and (4) Monetary policy rule.

Apart from being used in forecasting, the main goal of the MPT is simulating the response of the main macroeconomic variables (inflation, GDP gap, monetary policy interest rate and exchange rate) to different shocks. The MPT serves as a great tool for policymakers at the BCRP. However, the great number of equations involving it makes it a non-parsimonious model, as well as non-replicable for independent analysts and agents outside the projection process.

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 concerning inflation forecasting using ML models has mainly 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. On the other hand, 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 developing country 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). They used five different supervised ML algorithms: ridge regression, LASSO regression, EN, RF, and quantile regression forest (QRF), as well as a series of traditional econometric models: RW, VAR, and ARMA. Additionally, they employed reduced-form structural models, factor models, and survey forecasts. The survey forecast was used as 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. ML forecasts showed superior predictive power and outperformed the inflation forecasts of all traditional econometric approaches.

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). They used 102 monthly predictors, covering production, government debt, price indexes, taxes, financial markets, import and export of goods and services, government accounts, savings, investment, wages, and international variables recovered from the Central Bank of Brazil (BCB), other government databases (the FGV, the IBGE, the IPEADATA) and the Bloomberg database. The selected period ranged from January 2000 to December 2013, after the implementation of the inflation-targeting policy in Brazil. The predictions were made using a rolling window.

The models used in the paper were the autoregressive model (AR), the factor model, the LASSO regression, and the adaLASSO specifications for all the forecast horizons, which ranged from one to twelve months (h = 1, ..., 12). The metrics used to compare the different methods were the root mean squared error (RMSE) and the mean absolute error (MAE). 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). By using data from the Central Bank of Costa Rica (BCCR) from January 2003 to February 2019, he tested a univariate K-nearest neighbors (KNN) model, a KNN with explanatory variables, extreme gradient boosting, an RF model, and long short-term memory (LSTM) model.

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 the next section, we will describe the theoretical framework that is going to be used in this document for the in the evaluation and comparison of the forecasts. Then, we are going to present the methodology as well as the data and the different models for modelling inflation, both econometric and machine learning models.

3 Forecast Evaluation Framework

In this section, we will describe the criteria of forecast evaluation and forecast comparison. In the section 3.1, we will describe the two main properties of a good forecast: unbiasedness and efficiency. In the section 3.2, we will develop the main procedures when comparing forecasts.

Given a $X_t$vector of k explanatory variables for the variable y_t for $(t = 1, ..., T)$ we can construct the point forecast of y_{t+h} given the information I_h.

$\hat{y}_{t+h} = f_{t+h}(X_{t+h}, I_h)$

we define the forecast error as

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

3.1 Forecast evaluation

Following Granger & Newbold (1986) and Clements & Henry (1998) on forecast evaluation, we identify unbiasedness and efficiency as the two main properties of a good prediction.

3.1.1 Unbiasedness

Unbiasedness can be interpreted as the fact that the optimal forecast under a MSFE loss function is the conditional expectation of the variable.

We first introduce the idea of a cost (loss) function given the size of the error $C(e)$, then as a natural criterion, given such loss function, we will choose the point forecast that minimizes it:

$$f_{t+h} = minC(e)$$

Then, an unbiased forecast fulfills

$$f_{t+h} = E_c(X_{t+h}|I_h)$$

or, equivalently

$$E_t(y_{t+h}) = \hat{y}_{t+h|t} = f_{t+h}$$

where $E_t(.)$ is the conditional expectation given information at time t.

The result of the expected forecast error should be equal to zero, which means that, on average, the forecast should be correct.

To test for unbiasedness, we should consider the regression

$$\hat{y} = \alpha + \beta \hat{y}_{t+h} + \epsilon_{t+h} \tag{1}$$

for $t = T, ..., T + H - h$ where h is the forecast horizon and $T, ..., T + H - h$ is the evaluation sample.

For the forecasts$\hat{y}_{t+h}$\$ to be unbiased we can use a robust F-test to test the condition $\alpha = 0, \beta = 1$. We can also write the following equation:

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} = \epsilon_{t+h} + \tau \tag{2}$$

and test for $\tau = 0$ with a robust t-test for $\hat{y}_{t+h}$ to be unbiased.

3.1.2 Efficiency

This criterion refers to the efficient use of all the available information. If the forecast is inefficient, then it can be improved upon with a better specification in the forecasting model. We can test weak or strong efficiency. Weak efficiency means that the optimal h-steps forecast error (e_{t+h}) should be correlated, across time, at most of order $h-1$ and should be uncorrelated with available information at the time the forecast was made. This property can be evaluated by fitting a moving-average of order \$h-1\$ to the h-steps ahead forecast error (e_{t+h}) and then testing if the residuals are white noise.

Strong efficiency can be easily assessed by testing $\gamma = 0$ in

$$e_{t+h} = \gamma' z_t + \epsilon_{t+h} \tag{3}$$

where z_t are potentially relevant variables for the explanation of forecast errors. If the $h-$steps ahead forecast errors (e_{t+h}) are strongly efficient, then no indicators z_t at the time the forecast was formulated can improve them and, therefore, explain the $h-$step forecast error.

3.2 Forecast comparison

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 will 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 will outline a forecast comparison test that evaluates whether the difference between forecasts is statistically significant, namely the Diebold-Mariano test.

3.2.1 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

$$RMSFE = \sqrt{MSFE} = \sqrt{\frac{1}{T} \sum_{t=1}^{N} e_{t+h}^2} \tag{4}$$

3.2.2 Mean absolute percentage error (MAPE)

The mean absolute percentage error is a loss function defined as

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

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}|\frac{e_{t+h}}{y_{t+h}}| \tag{5}$$

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

$$DM = H^{1/2}\frac{\sum_{j=1}^{H} d_j/H}{\sigma_d} = H^{1/2}\frac{\bar{d}}{\sigma_d} \tag{6}$$

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 varianceγ_0 can be used.

4 Methodology and data

In this section we will first present the data and discuss the models that are going to be compared as well as the methodology to assess the models' performance. There are two main periods to be considered for the forecast evaluation. The first one ranges from 2008M1-2008M12, which was a period where energy and commodity prices, especially oil prices, boosted inflationary pressure, which greatly impacted both advanced and emerging economies (International Monetary Fund, 2008). The second period ranges from 2019M1 to 2019M12, were the was a small increase in local inflation due to foodstuffs and electricity rates, as well as major volatility (BCRP, 2015). Forecasts will we done in the short horizon, ranging from from a 3-months period, 6-month period and a year ($h = 3, 6, 12$). The third period, the most recent one, tries to predict inflation from 2023M1 to 2023M12.

Models will we rank based on their RMSFE and their MAPE for each period in all three horizons. We will then use the Diebold-Mariano test to compare each model against the benchmark RW and see how each model competes the benchmark RW.

We are considering the two econometric models (RW, VAR) and four machine learning models (LASSO, Ridge, EN and RF).

4.1 Data

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.

4.2 Econometric models

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

4.2.1 Random Walk (RW)

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

$$y_t = y_{t-1} + \epsilon_t \tag{7}$$

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

$$\hat{y}_{T+h} = y_T \tag{8}$$

The forecast errors are

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

Hence, the variance of the forecast errors are

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

4.2.2 Vector Autoregression (VAR)

Given a VAR(p) model of k variables$y_{1t}, ..., y_{kt}$ grouped in a $(k \times 1)$ vector $y_t = (y_1t, ..., y_kt)'$, it can be represented as:

$$y_t = \mu + \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + \epsilon_t \tag{9}$$

$$\epsilon_t \sim WN(0, \Sigma) \tag{10}$$

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^2p 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(∞)

$$y_t = \Psi(L)\epsilon_t \tag{11}$$

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

$$\hat{y}_{T+h} = \sum_{j=0}^{\infty} \Psi_{j+h} \epsilon_{T-j} \tag{12}$$

with the following forecast error

$$e_{t+h} = \sum_{j=0}^{h-1} \Psi_j \epsilon_{t+H-j} \tag{13}$$

and the variance-covariance of the forecast error

$$V(e_{t+h}) = \Sigma + \Psi_1 \Sigma \Psi_1' + ... + \Psi_{h-1} \Sigma \Psi_{h-1}' \tag{14}$$

4.3 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.

4.3.1 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 linear regression model

$$y_t = \beta x_t + \epsilon_t$$

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 vector regression coefficients and ϵ is a vector of errors. It is possible that K is relatively large compared to N. In those cases, the LASSO estimates are chosen to minimize

$$LASSO = \min_{\beta}(\sum_{i=1}^{N}(y_i - \sum_{j=1}^{K}\beta_j x_i j)^2 + \lambda \sum_{i=1}^{K}|\beta_i|) \tag{15}$$

$$LASSO = \min_{\beta}(RSS + \lambda \sum_{i=1}^{K}|\beta_i|) \tag{16}$$

where RSS is the residual sum of squares and the second term $\lambda \sum_{i=1}^{K}|\beta_i|$ is called the shrinkage penalty and is a regularization type ℓ_1. 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. Following the literature, $\lambda = 0.5$ for this document. LASSO yields sparse models, which are models that involve only a subset of the variables.

The process of reducing variables is called feature selection. Once it is completed, we realize a k-fold cross-validation that splits the data into test and training samples. Each iteration, the k-fold is used as the testing set, and the remaining k-1 folds are used as the training set. This means that the ML model is trained k times, each time using a different fold. Upon completion, we proceed to do the point forecasts.

4.3.2 Ridge Regression

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

$$Ridge = \min_{\beta}(RSS + \lambda \sum_{i=1}^{K}\beta_i^2) \tag{17}$$

where the term $\lambda \sum_{i=1}^{K}\beta_i^2$ is a regularization of typeℓ_2 and λ is the tuning parameter.

As λ increases, the coefficients $\beta_1, ..., \beta_k$ will approach zero. λ will be set to 0.5 in this document. 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 with the ℓ_2 penalty, the k-fold cross validation and the train-test split begins. The ML is trained k times and then we start with the forecast procedure.

4.3.3 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

$$Elastic\ Net = \min_{\beta}(RSS + \lambda(\alpha \sum_{i=1}^{K} |\beta_i| + (1-\alpha) \sum_{i=1}^{K} \beta_i^2)) \tag{18}$$

$$Elastic\ Net = \min_{\beta}(RSS + \lambda(\alpha\ell_1 + (1-\alpha)\ell_2) \tag{19}$$

where ℓ_1 and ℓ_2 correspond to their specific regularization and α 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 to a LASSO regression. By convention, we will use $\alpha = 0.5$ which will give an equal proportion to each regularization technique. The other tuning parameter λ presents the weight of the combined penalties. It is also set to 0.5.

After the selection of the variables, the k-fold cross-validation starts the train-test split is executed. Once the ML model has been trained k times, we can start the forecasting procedure.

4.3.4 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

$$y = \sum_{i=1}^{p} X_i\beta_i \tag{20}$$

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

$$y = \sum_{m=1}^{M} c_m \cdot 1_{(x\in R_m)} \tag{21}$$

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 (30) 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 (31) 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.

4.4 Model implementation

All models are implemented using the Scikit-Learn and XGBoost package in Python. Linear models are imported as the Lasso, Ridge, and ElasticNet functions respectively. Non-linear models are imported from RandomForestRegressor and XGBRegressor. All models are implemented with a random state = 2023. A cross validation followed by a grid-search is implemented using the TimeSeriesSplit and GridSearchCV modules from Scikit-Learn. The models best parameters using the grid-search are as follows:

- Ridge: $\lambda = 0.1$
- Lasso: $\lambda = 0.1$
- Elastic Net: $\lambda = 0.1, \alpha = 0.1$
- Random Forest: $maxdepth = 10, nestimators = 150$

5 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 two periods January 2009 to December 2009 and January 2019 to December 2019. Both metrics are presented using the RW as the benchmark. In order to test for robustness, we are making our predictions for two variables: Headline Inflation (Lima Metropolitan Precious Index (monthly % change) - CPI) and Core Inflation (Metropolitan Lima price index (monthly % change) - Underlying CPI). We are using the first and second lags of the other variables as predictors. We will first present the results for the Headline Inflation as the target variables in both 2009 and 2019. Then, we will present results for the Core Inflation in the same periods.

5.1 2019

5.1.1 Forecasting from January 2009 to December 2009

We can see in figure A3 the results for the RMSE. It is noticeable that for all models it lies under 1, implying all models presented (linear and non-linear) are improving upon the RW prediction. There are three groups clearly identified. First, we observe that the RF model has the best performance of all models for all horizons (Except for a tie on $h = 2$ with the EN model). There is even a slight decrease in the RMSE after the second horizon, unlike other models. Second, we see all linear ML models competing for the second place. Finally, the VAR model has the worst performance overall, with an increasing RMSE from around 0.3 in $h = 1$ to 0.5 in $h = 12$.

In figure A4 we have plotted the MAPE for all models as a percentage of the RW RMSE. We see that the RF has overall an absolute error of less than 10%. It is followed closely in the nearest horizons by the EN and Lasso linear ML models, which have almost the same results up to $h = 8$. The Lasso model performs better than the other linear models only in the most distant horizons

($h = 8$ and higher). It also should be noted that the MAPE is higher in both the shorter and longer horizons, following a similar pattern for all linear models.

Regarding the main predictors, we will discuss the ones elected by the best two models overall: the RF model and the EN model. The linear ML model is choosing typical macroeconomic variables related to inflation such as CPI Tradables,the Monetary Policy Reference Rate and the price of commodities such as Soybean Oil and Wheat. The RF model, on the other hand, chooses as its main predictor the real index of the minimum living wage (MLW), and then, with a lesser strength, the CPI Non-tradables and the CPI Food and Beverages. In Table A1 we can see all 10 predictors for both models.

5.1.2 Forecasting from January 2019 to December 2019

We observe the results for the RMSE for the second forecast in figure A5. Again we observe the RF as the best model across all horizons. It increases slightly in the second horizontal and then continues to decrease until it stabilizes at $h = 6$ onwards. The RF is followed closely by the linear ML models in second place. The VAR model falls far behind with the higher RMSE.

The MAPE for all models in this prediction in far greater than the 2009 forecast, exceeding 100% for all non-linear models. The RF has the best MAPE across all horizons with less than 50%. Both the linear ML and the VAR models follow the same pattern, with higher MAPE for shorter and longer horizons.

Regarding the predictors chosen, we are again centering in a linear and a non-linear ML model. EN chooses again typical variables used to predict inflation, such as the lags of the Monetary Policy Reference Rate and commodity prices, while RF opts for the lags of the index of the MLW. In Table A2 we can see all 10 predictors for both models.

5.2 Core inflation

5.2.1 Forecasting from January 2009 to December 2009

Figure A7 shows the RMSE for the predictions of core inflation from January 2009 to December 2009. The results are similar to those shown for headline inflation. RF outperforms all models for all horizons with the smaller RMSE. The linear ML models follow the same pattern with an increasing their RMSE as the horizon increases. The VAR model has the worst performance, event worse than the RW model.

In figure A8 we observe the results for the MAPE. Here we find results inversed from all previous cases. First, we observe that all models have a similar MAPE in the shortest horizons. In the longer horizons ($h = 6$ and onwards), the MAPE for the VAR is smaller than other models. The RF is the second best model in terms of MAPE. Finally, the linear ML models follow a similar pattern with the highest MAPE.

Regarding the most important covariates, the RF no longer chooses the MLW index as its predictor. Instead, it opts for different Underlying CPI indicators as well as commodity prices (Wheat, Soybean oil). The linear EN chooses traditional variables such as the lags of CPI and the reserve requirement rate. We can see all predictors in Table A3.

5.2.2 Forecasting from January 2019 to December 2019

Again results from both figure A9 show outcomes different from previous forecasts. It appears that the VAR model performs better at longer horizons, followed by the non-linear RF. This last model also has the worst performance during the first horizons. The Ridge model has the best RMSE at the first horizons but it get continuously worse as the horizons increase.

Regarding the results from the MAPE, similarly to the results from the RMSE, both the VAR and RF have a better performance, with the first one being the best of the two models at all horizons. The linear ML models behave following a similar pattern with an increasing worse MAPE over time.

When it comes to the predictors chosen by the best models, we center this time in the Ridge linear model and the RF model. Both the linear and the non-linear ML models choose typical variables used to model and predict inflation. This time, the RF model has no preference for the MLW index, and chooses instead underlying inflation, as well as soybean oil and wheat prices to predict core inflation. The complete list is in Table A4.

6 Conclusions

Our main findings in this thesis are similar to those presented in other researches, however, there are some unique findings regarding Peruvian inflation. First of all, ML clearly outperform the traditional VAR and RW models in terms of RMSE and MAPE when it comes to headline inflation prediction for all horizons. It is also evident that the non-linear RF model has a better RMSE and MAPE than linear ML models like Ridge, Lasso and EN. It is true for both testing periods, 2009 and 2019.

In both periods, the linear models followed a similar pattern, with a performance in- between the VAR and RF model. It is also noticeable that, out of all linear models, the EN had better results in both short and long term predictions, followed closely by the Lasso model.

However, results vary when forecasting core inflation. In the period of 2009, While the RF is still the best model when measuring RMSE for all horizons, surprisingly the VAR has a better MAPE than all models at the longest horizons ($h = 5$ and on-wards), leaving the RF at second place. Additionally, linear ML models have better MAPE results in the shortest horizons ($h = 1, 2, 3$). In the period of 2019, results completely change from those of headline inflation, because the VAR model has better results in both the RMSE and MAPE metrics, specially in longer horizons.

This could be interpreted as follows: headline inflation has more non-linearity's in the relationship between the variable and its covariates, while core inflation has more linear and traditionally-modelled relations. That explains why the VAR model fails to give better predictions for headline inflation, but has better results for the core inflation. This linear relationships are stronger in the shortest horizons, which explains why linear ML models have better RMSE and MAPE metrics up to $h = 5$, but then are outperformed by the RF model. It is worth mentioning that the non-linear RF still gives a fair fight, particularly in the longest horizons.

Finally, regarding the most important predictors chosen, we will focus and those chosen by the best two models in each period. For headline inflation, it is worth mentioning that, while the linear EN model chooses typical variables for modelling inflation, such as other CPI indicators, the reference rate and the lags of the prices of commodities, the RF model selects the MLW index as its main predictors in both periods (2009 and 2019). This results imply that, while traditional variables have a linear relationship with headline inflation, variables added to the model such as MLW have a non-linear non-direct relationship with it. For core inflation this results slightly change. In the period of 2009 both EN and RF select traditional variables as their main predictors, an the MLW index comes 10th for RF in terms of importance. In 2019 the same happens, but this time with the Ridge and RF model. It is also noticeable that the Ridge model chooses the lags of liquidity, which is something that other models did not choose.

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] Jairo, F., & Rodrigo, G. (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

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

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

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

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

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

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

[19] 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.

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

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

[22] 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

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

[24] 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

[25] 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.

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

[27] 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

[28] 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

[29] 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

[30] Morgentsten, O. (1928). *Wirtschaftsprognose: eine Untersuchung ihrer Voraussetzungen und Möglichkiten.*Vi Julius Springer.

[31] Pincheira-Brown, P., Selaive, J., \& Nolazco, J. L. (2019). Forecasting inflationhttps://doi.org/10.1016/s0169-2070(01)00156-x in Latin America with core measures. International Journal of Forecasting, 35(3), 1060–1071. https://doi.org/10.1016/j.ijforecast.2019.04.011

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

[33] 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

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

[35] 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.

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

[37] 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

[38] 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

[39] 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.

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

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

[42] 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

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

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

[45] 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*.

[46] 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

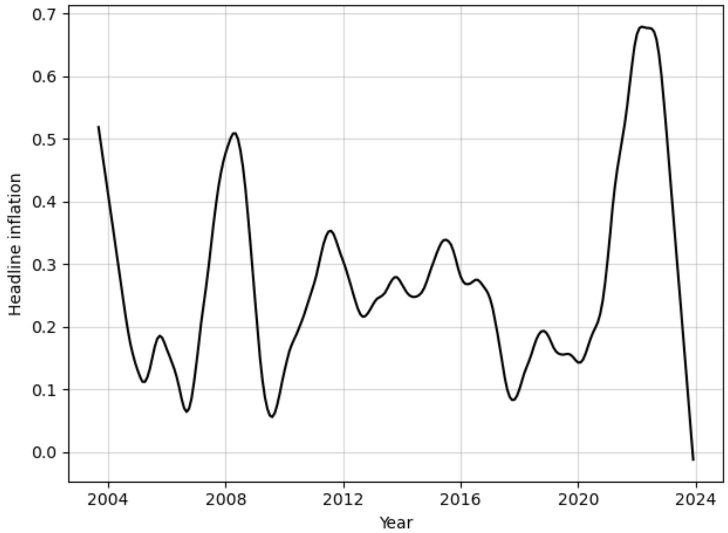
[47] Vega, M., Bigio, S., Florián, D., Llosa, G., Miller, S., Ramírez, N., Rodríguez, D., Salas, J., & Winkelried, D. (2009). Un Modelo Semi-estructural de Proyección para la Economía Peruana.*Revista Estudios Económicos, (19).* https://www.bcrp.gob.pe/docs/Publicaciones/Revista-Estudios-Economicos/17/Estudios-Economicos-17-2.pdf

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

[49] 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.

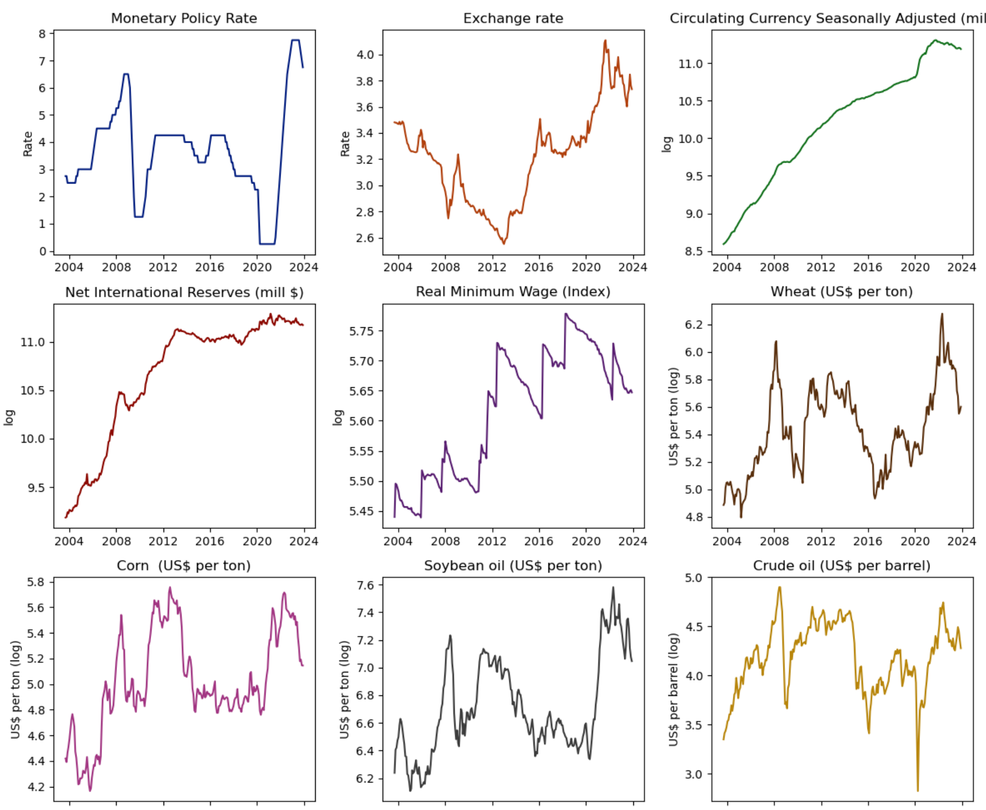
A Appendix

Figure A1: Headline Inflation



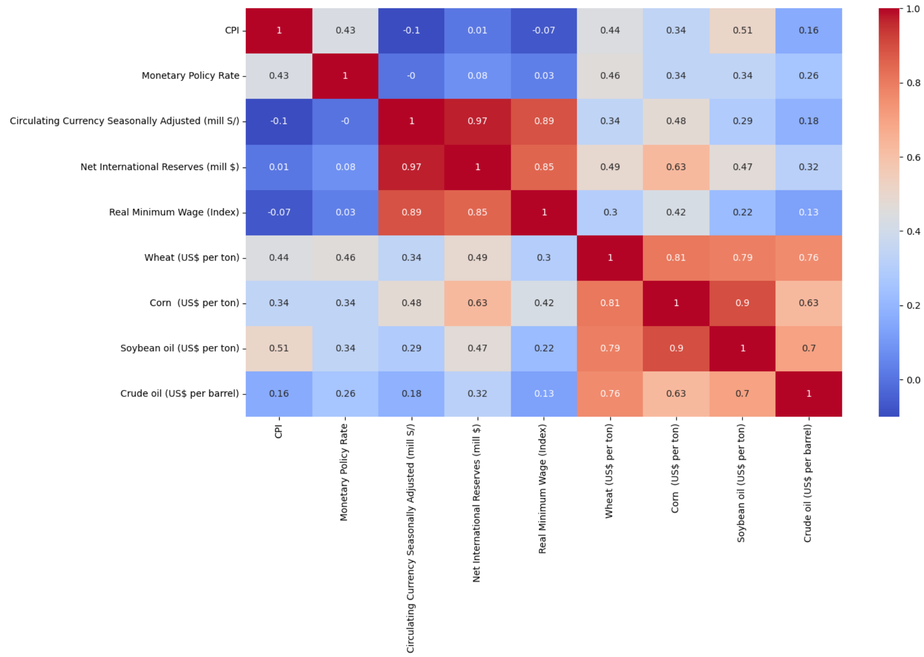
Headline Inflation (2003M9-2023M12)

Figure A2: Predictors for Headline Inflation



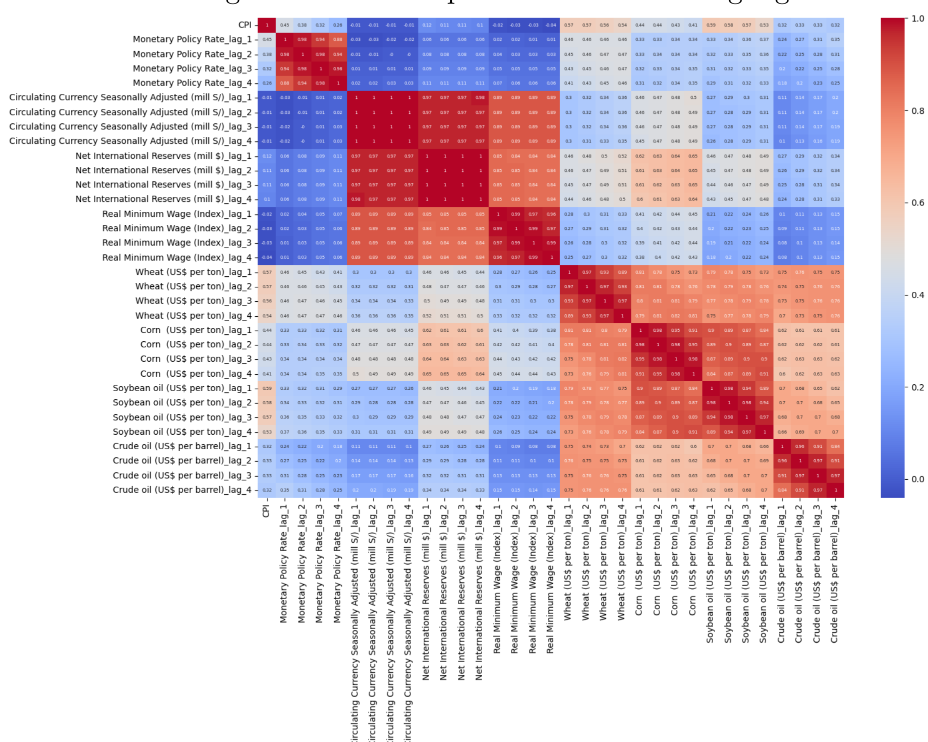
Variables used for the prediction of Headline Inflation (2003M9-2023M12)

Figure A3: Heatmap of variables



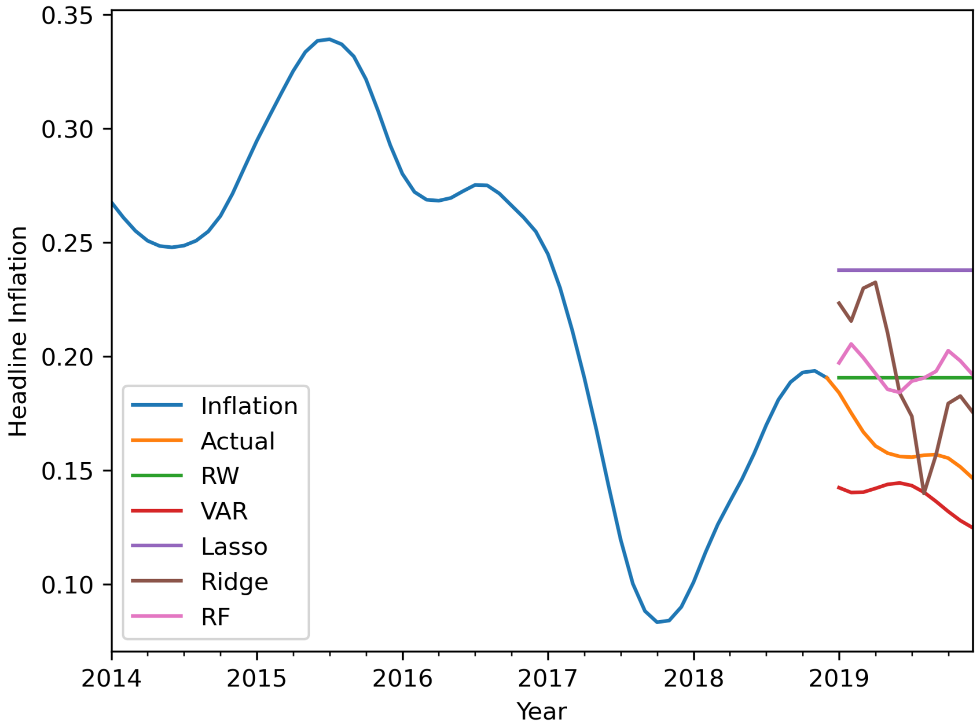
Heatmap correlation of headline inflation with all variables

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 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

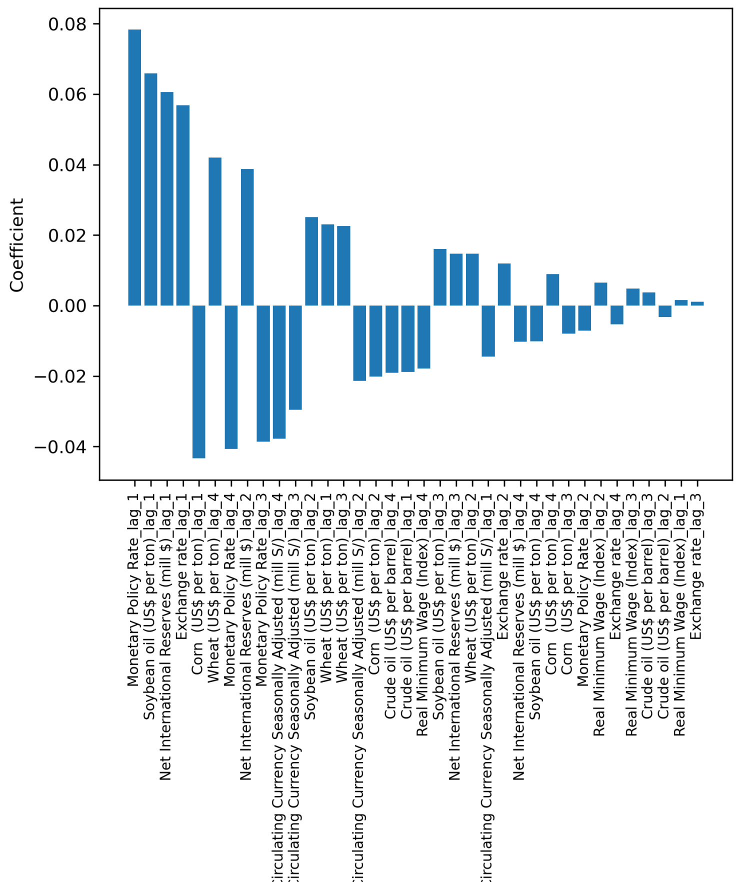
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model is outlined in bold.

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

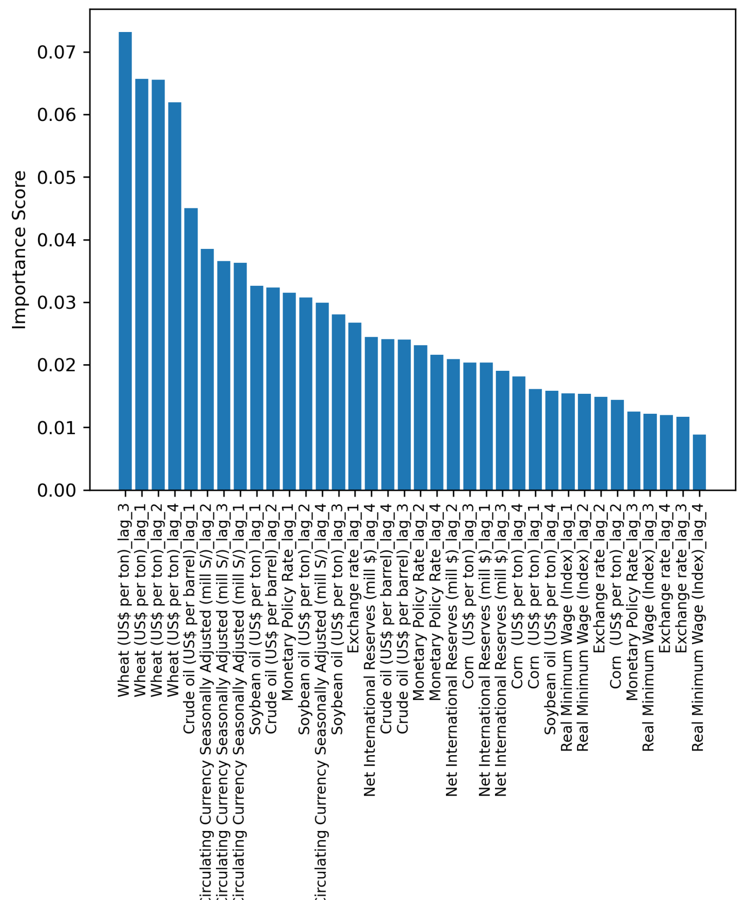
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A6: Main predictors of ridge regression



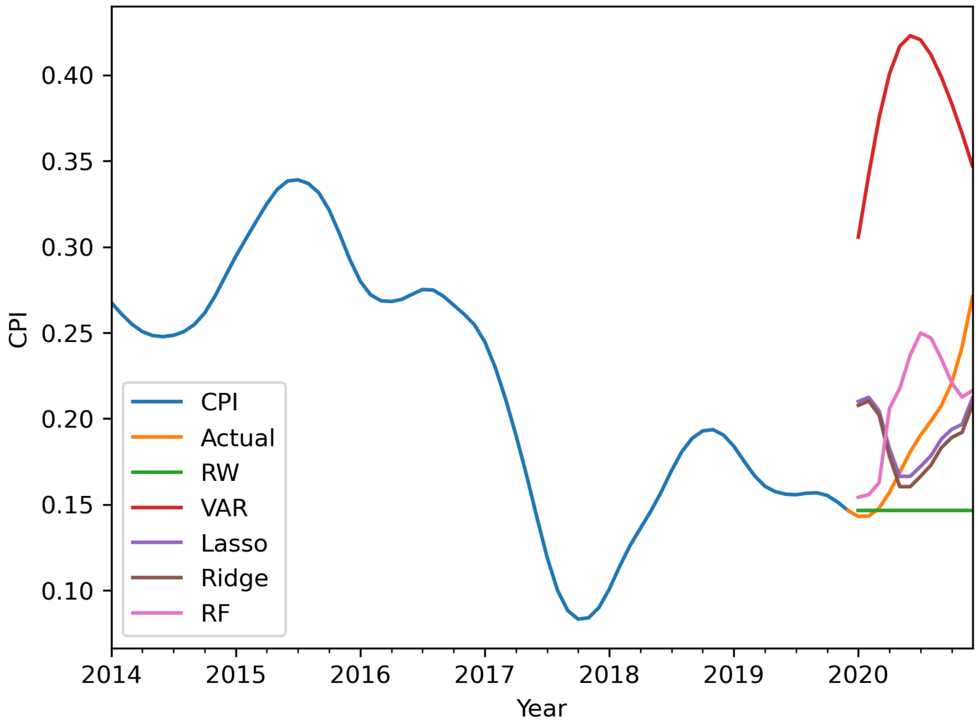
Main predictors chosen by the Ridge regression model for the 2019 forecast.

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 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**

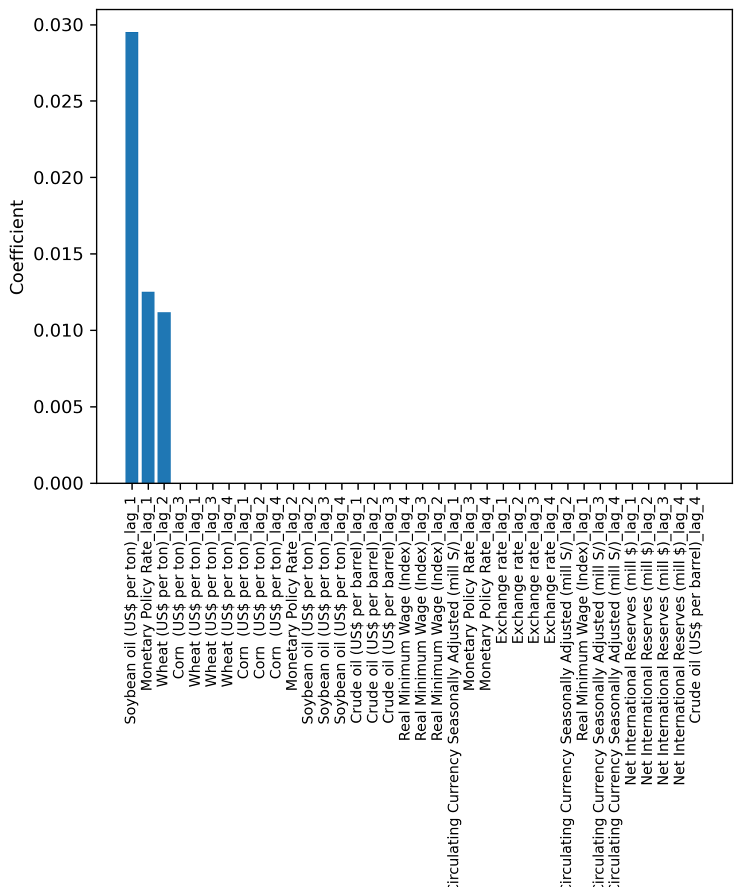
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model is outlined in bold.

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

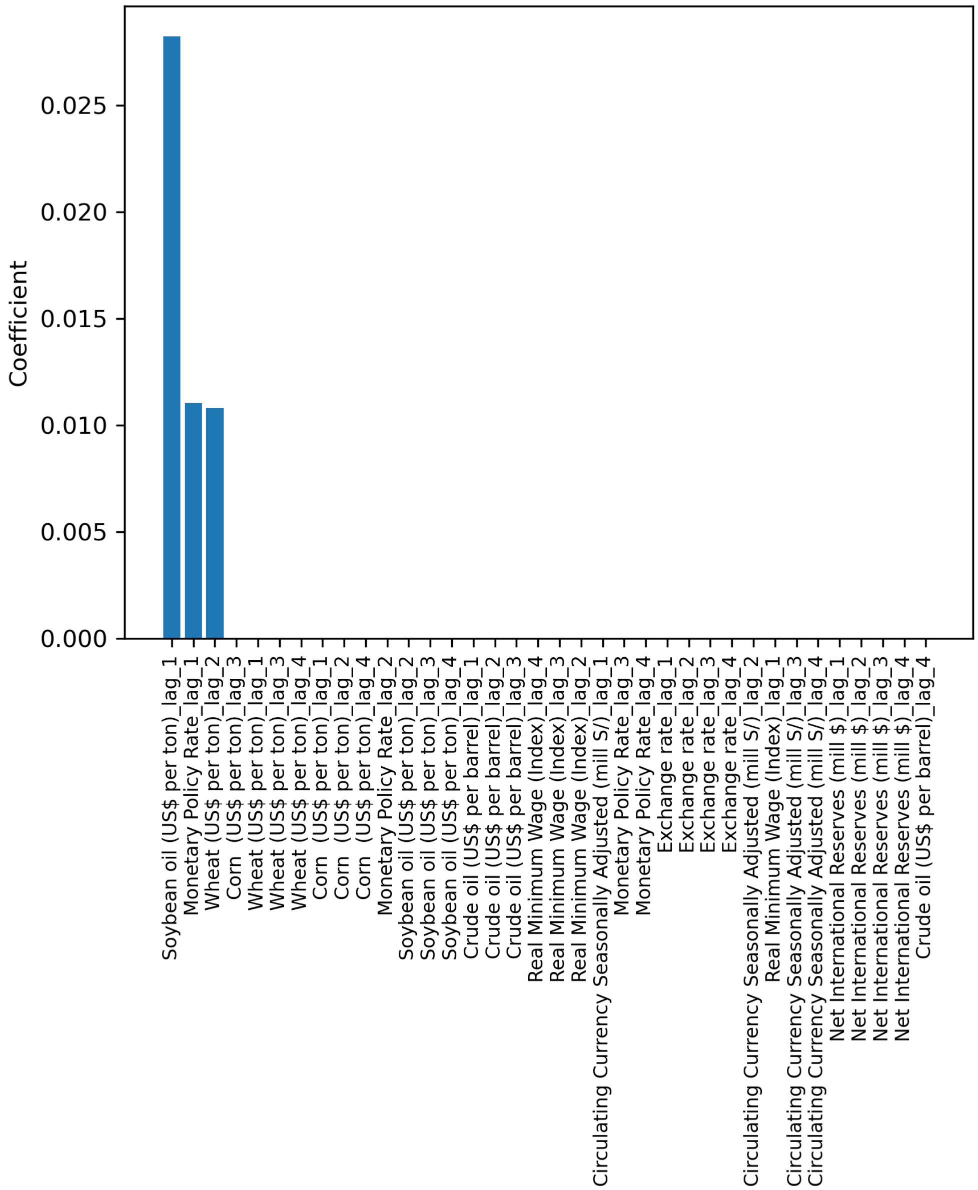
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A9: Main predictors of ridge regression



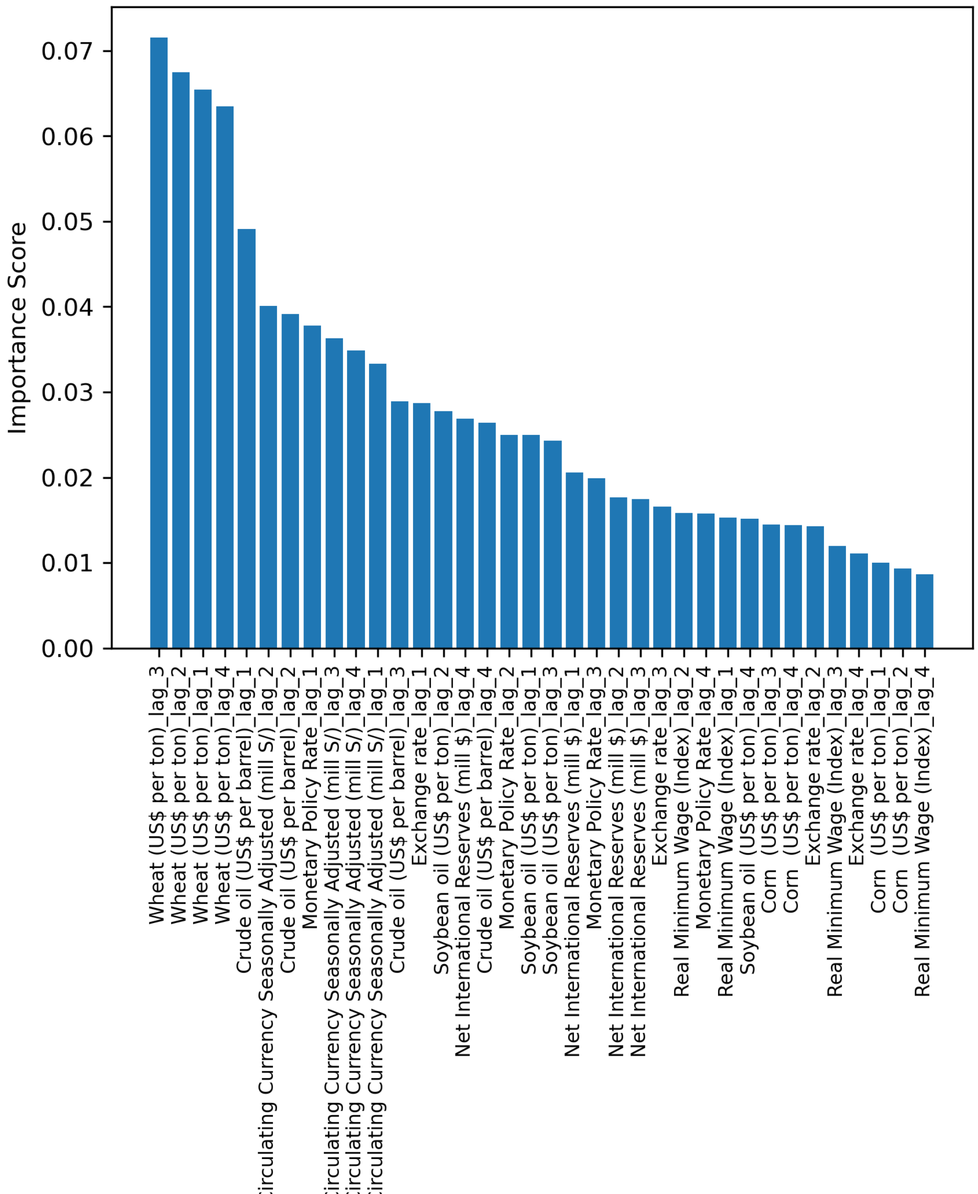
Main predictors chosen by the ridge regression model for the 2020 forecast.

Figure A10: Main predictors of LASSO regression



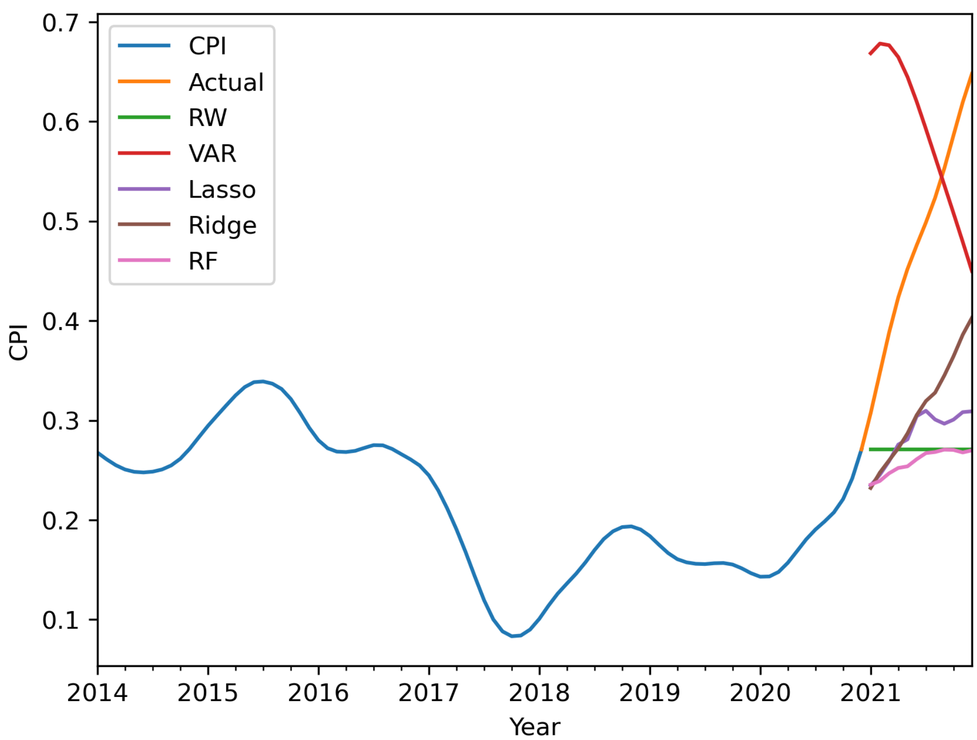
Main predictors chosen by the LASSO regression model for the 2020 forecast.

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												
VAR												
Ridge												
Lasso												
RF												

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 A8: MAPE of forecasts (2021)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark												
VAR												
Ridge												
Lasso												
RF												

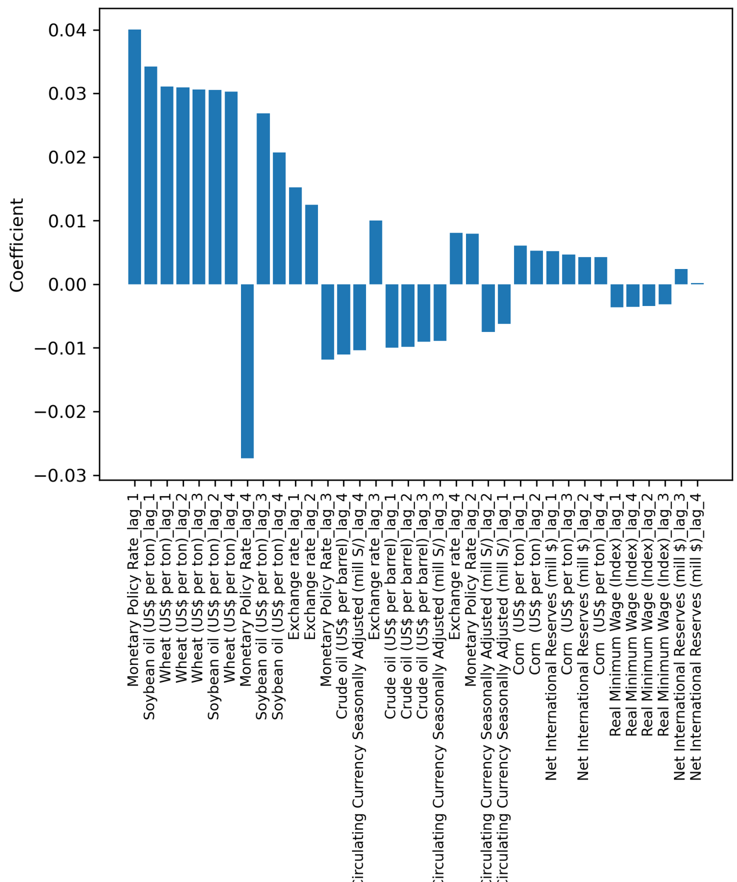
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model is outlined in bold.

Table A9: Diebold-Mariano test for 2021 forecast

Model	DM	p-value
VAR		
Ridge		
Lasso		
RF		

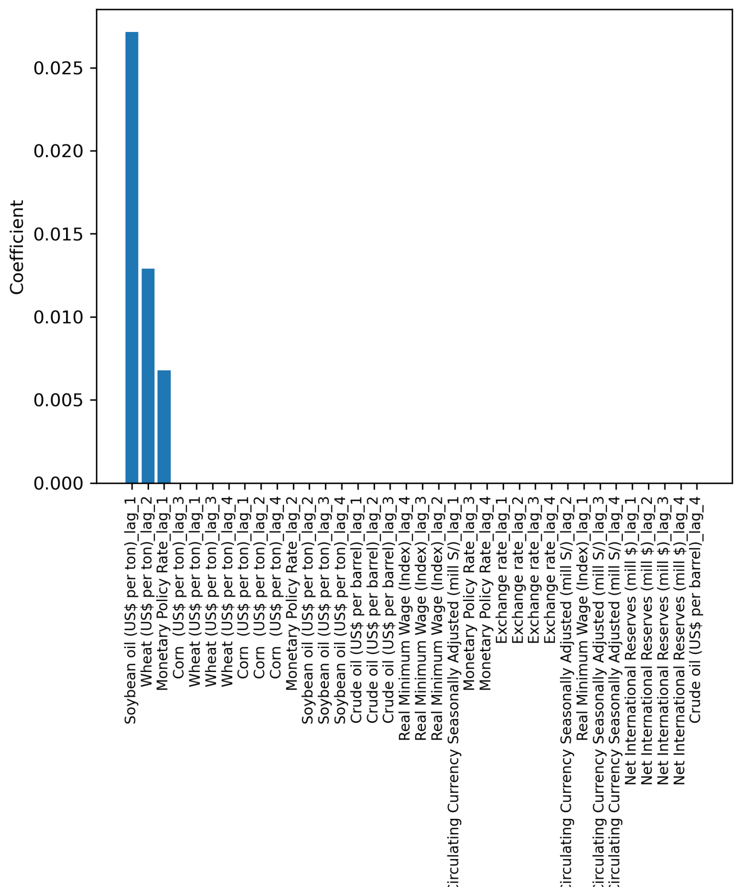
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A13: Main predictors of ridge regression



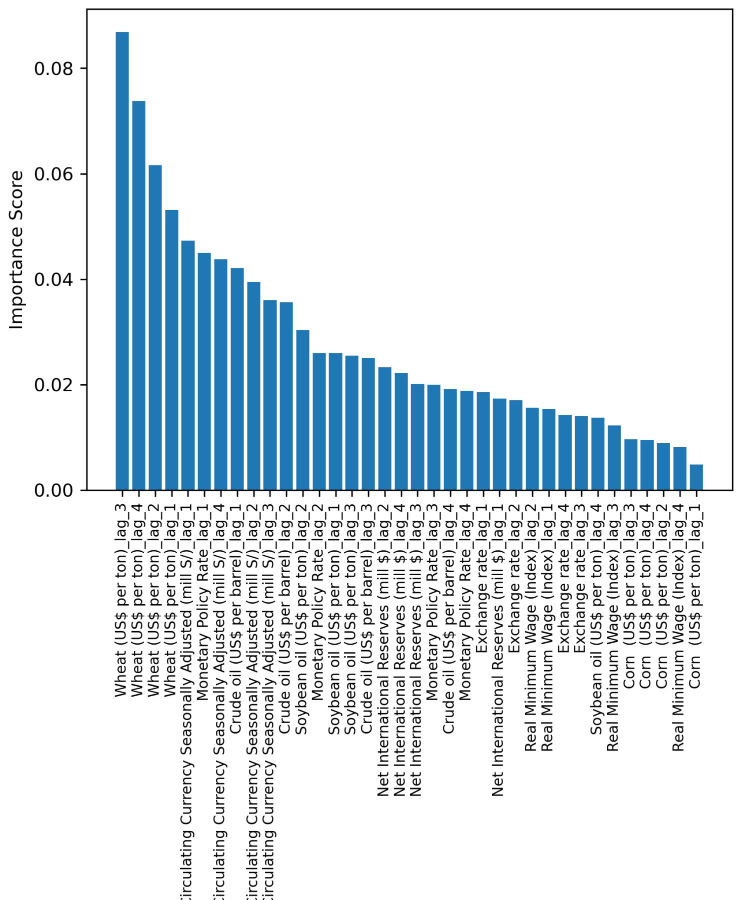
Main predictors chosen by the ridge regression model for the 2021 forecast.

Figure A14: Main predictors of LASSO regression



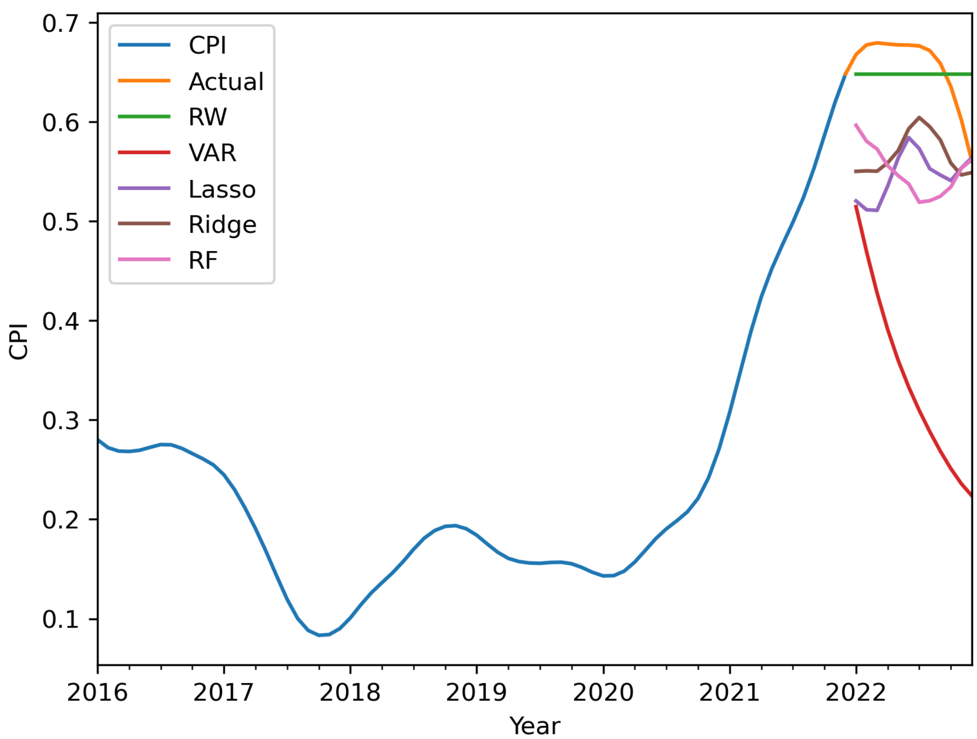
Main predictors chosen by the LASSO regression model for the 2021 forecast.

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												
VAR												
Ridge												
Lasso												
RF												

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 A11: MAPE of forecasts (2022)

horizon	1	2	3	4	5	6	7	8	9	10	11	12
Benchmark												
VAR												
Ridge												
Lasso												
RF												

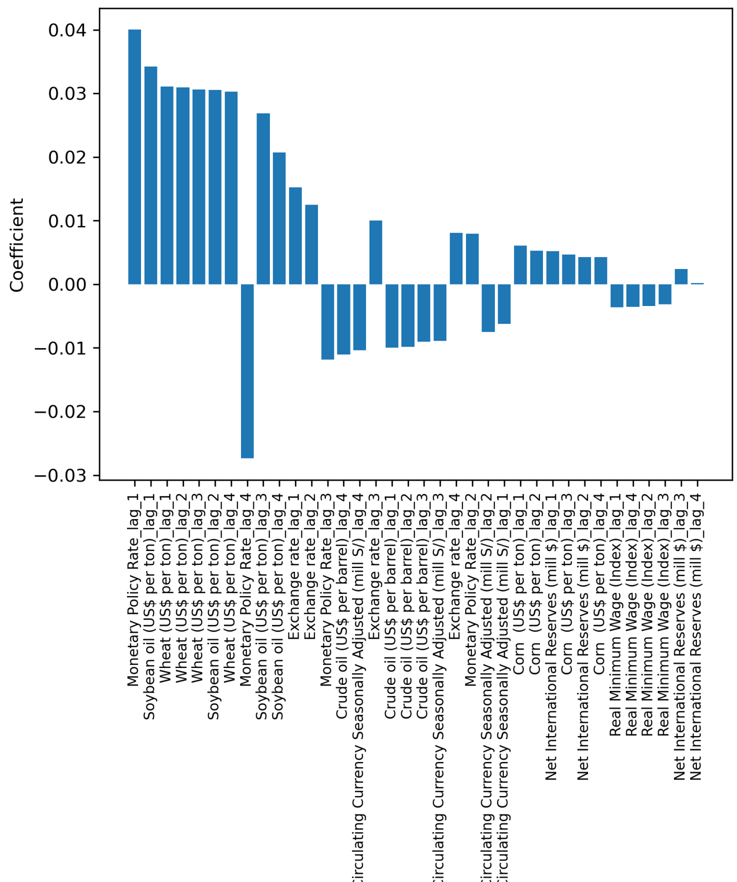
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model is outlined in bold.

Table A12: Diebold-Mariano test for 2022 forecast

Model	DM	p-value
VAR		
Ridge		
Lasso		
RF		

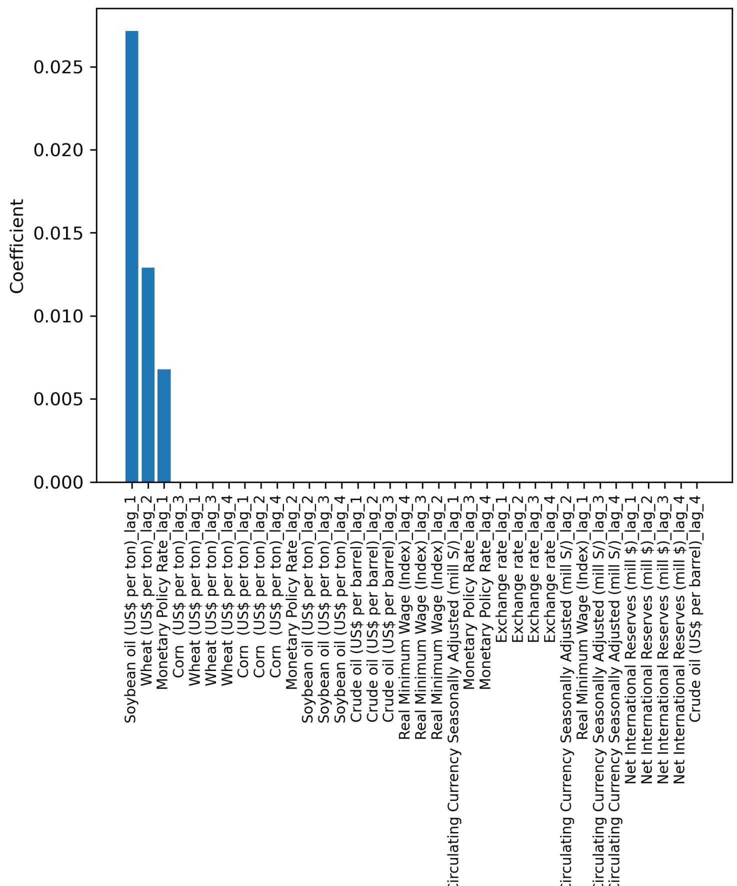
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A17: Main predictors of ridge regression



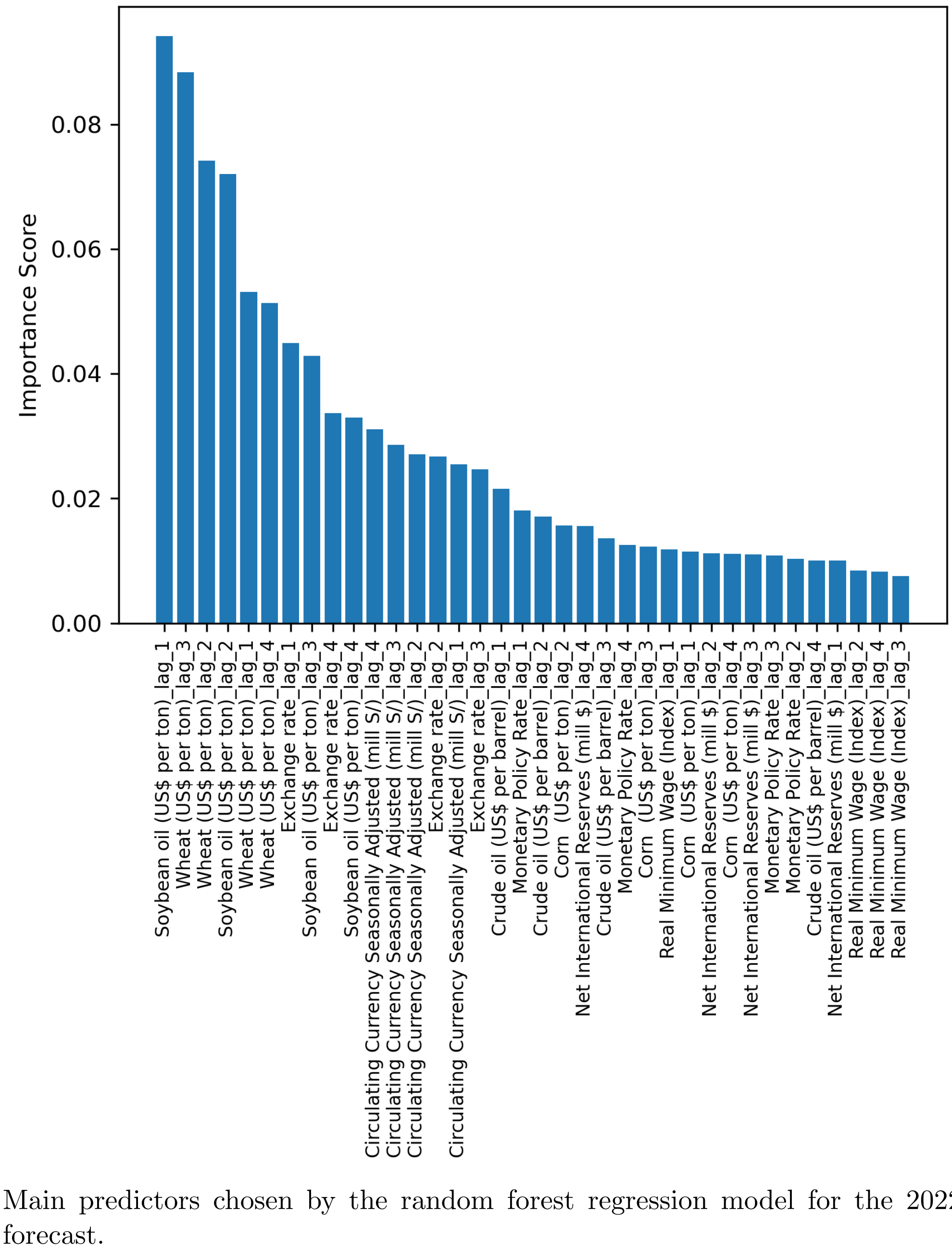
Main predictors chosen by the ridge regression model for the 2022 forecast.

Figure A18: Main predictors of LASSO regression



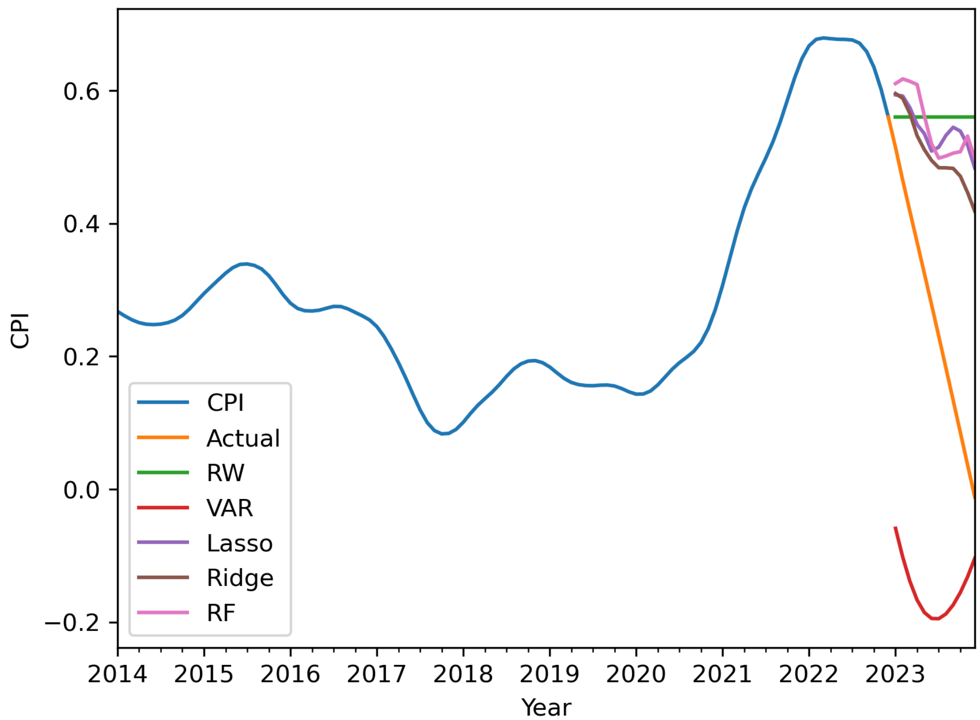
Main predictors chosen by the LASSO regression model for the 2022 forecast.

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												
VAR												
Ridge												
Lasso												
RF												

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												
VAR												
Ridge												
Lasso												
RF												

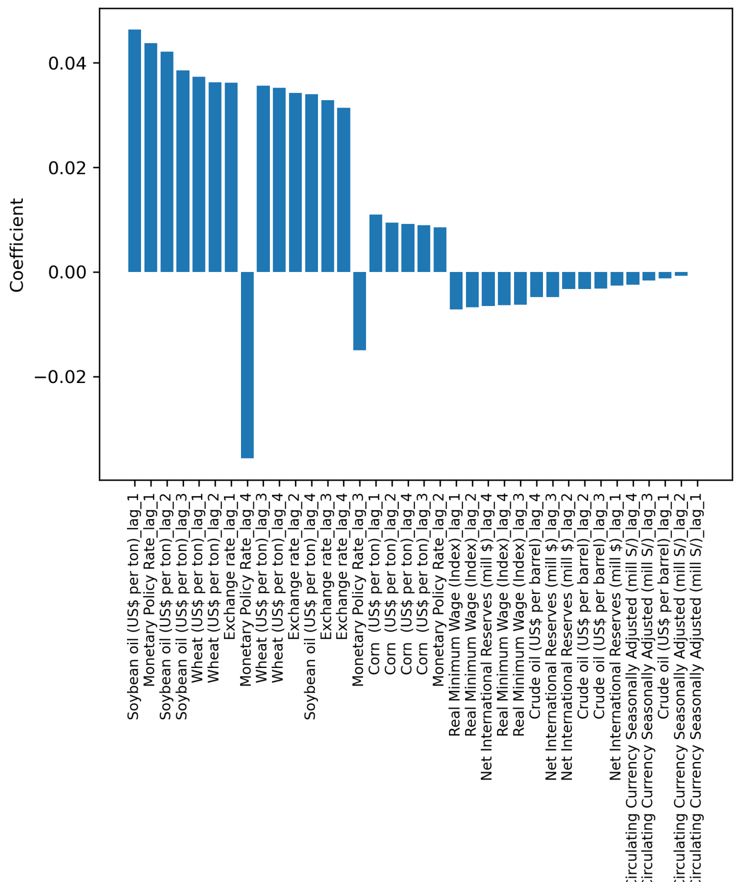
MAPE for Headline Inflation forecasts. The MAPE of other models is divided by the MAPE of the benchmark RW model. The best model is outlined in bold.

Table A15: Diebold-Mariano test for 2023 forecast

Model	DM	p-value
VAR		
Ridge		
Lasso		
RF		

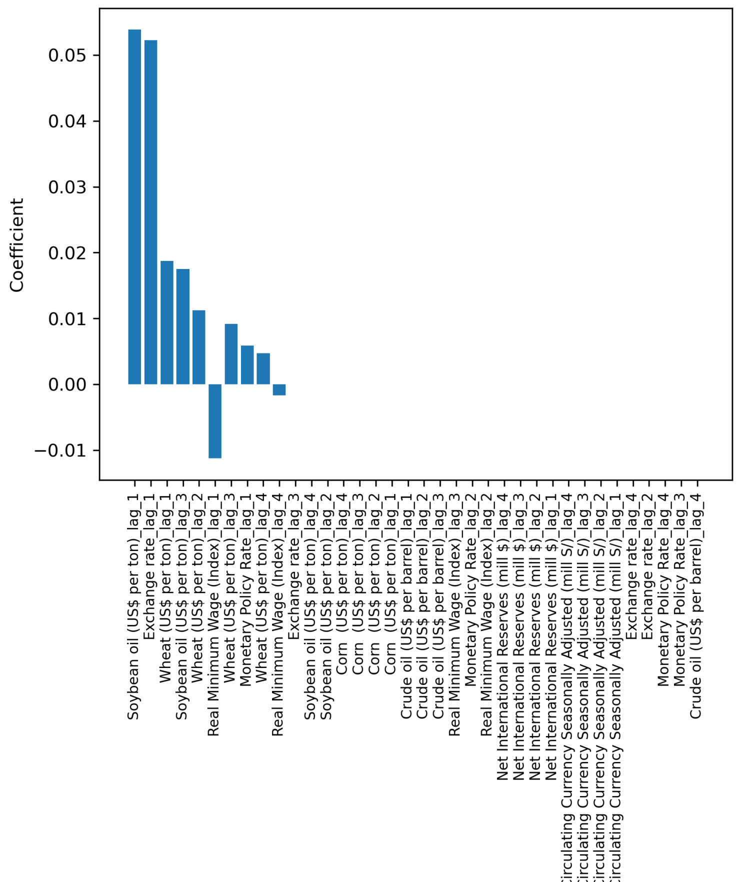
Results of the Diebold-Mariano test. We are testing all forecasts againts the RW forecast.

Figure A21: Main predictors of ridge regression



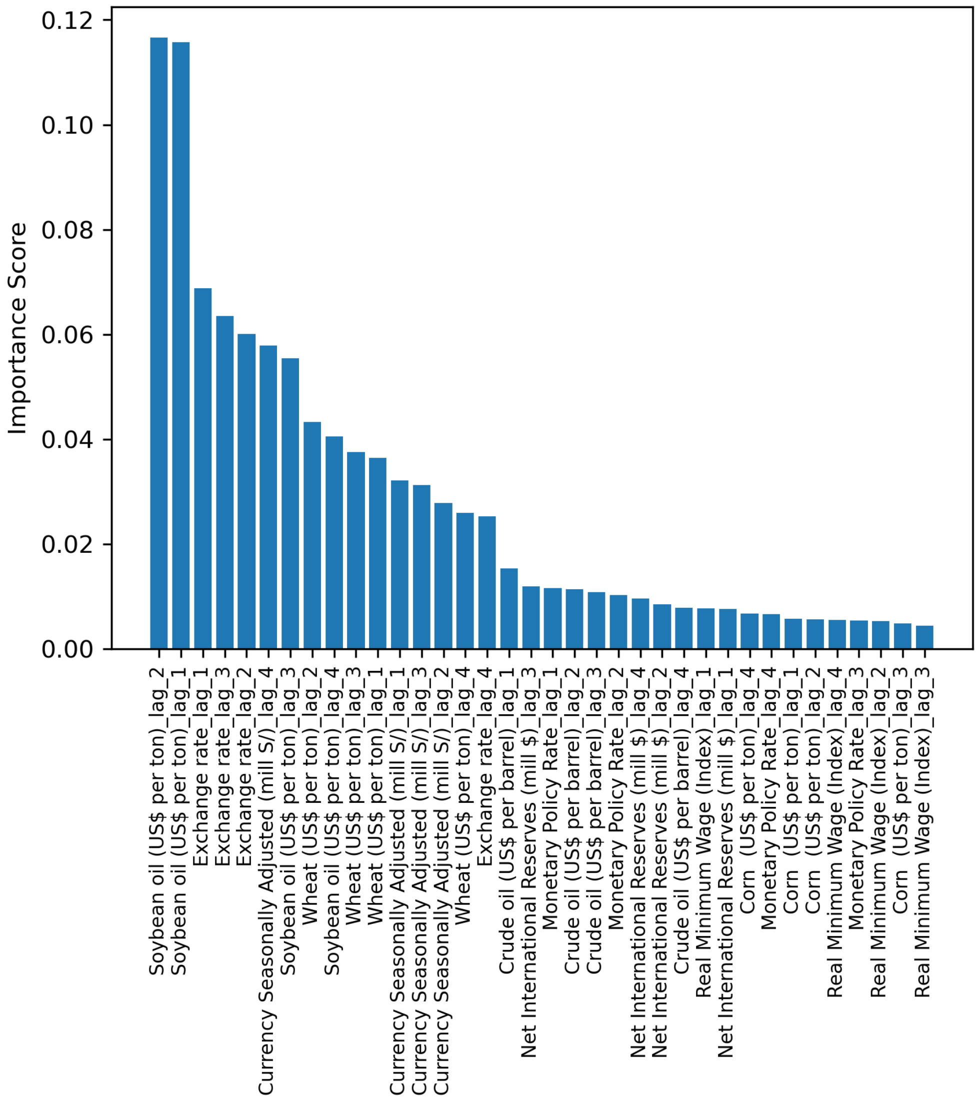
Main predictors chosen by the ridge regression model for the 2023 forecast.

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 2022 forecast.