Explainable Inflation Forecasts by Machine Learning Models

Serkan Aras[1,*], Paulo J.G. Lisboa[2]

[1] Econometrics Department, Faculty of Economics and Administrative Sciences, Dokuz Eylul University, Izmir, Turkey.
[2] Department of Applied Mathematics, Liverpool John Moores University, 3 Byrom Street, Liverpool, L3 3AF, UK.
* Correspondence: serkan.aras@deu.edu.tr; Tel.: +90-555-254-37-28

Abstract: Forecasting inflation accurately in a data-rich environment is a challenging task and an active research field which still contains various unanswered methodological questions. One of them is how to find and extract the information with the most predictive power for a variable of interest when there are many highly correlated predictors, as in the inflation forecasting problem. Traditionally, factor models have been used to tackle this problem. However, a few recent studies have revealed that machine learning (ML) models such as random forests may offer some valuable solutions to the problem. This study encourages greater use of ML models with or without factor models by replacing the functional form of the forecast equation in a factor model with ML models or directly employing them with several feature selection techniques. This study adds new tree-based models to the analysis in the light of the recent findings in the literature. Moreover, it proposes the integration of feature selection techniques with Shapley values to find out concise explanations of the inflation predictions. The results obtained by a comprehensive set of experiments in an emerging country, Turkey, facing a high degree of volatility and uncertainty, indicate that tree-based ensemble models can be advantageous by providing better accuracy together with explainable predictions.

Keywords: Machine learning; Inflation Forecasting, Factor Models, Model Interpretability, Tree-based Models; Shapley Values.

1. Introduction

Forecasting critical macroeconomic variables as precisely as possible is of great importance for authorities and decision-makers to make their economic policy more effective and powerful. This will positively contribute to confidence in the economy by decreasing uncertainty on the key economic indicators. Emerging economies like Turkey, in which general price levels are constantly high and very volatile, suffer more from high inflation rates than advanced economies since it shortens the duration of investment projects and substantially decreases saving and real incomes. Therefore, it is an uphill battle and of paramount importance for an emerging economy to mitigate the detrimental effects of inflation by forecasting it precisely. The task of making the economy more stable is generally undertaken by central banks all around the world. The difficulty of forecasting macroeconomic indicators stems from the presence of many highly correlated variables that affect the concerned variable to be forecasted. Hence, how a forecasting model is constructed is an important research problem when there is a large pool of candidate correlated variables.

To deal with this problem, factor models have been utilised to summarise the information on a big dataset from which the factors are extracted. Thanks to factor models, instead of putting all variables into a forecasting equation, only a few factors are modelled. However, it is not certain whether using all independent variables at hand or a reduced set of them via pre-selection algorithms will provide more informative extracted factors. Some studies claim that factor models based on a higher number of variables will not lead to better forecasting performance as compared to the factor models relying on a small number of variables (Alvarez et al., 2012; Boivin & Ng, 2006). Also, the selection of the number of factors to be used in a forecasting equation is still another unanswered problem (Agostino & Giannone, 2012; Schumacher, 2007).

An alternative approach to forecast an economic variable in data-rich environments is to take advantage of variable selection techniques in reducing the number of covariates. Variable selection techniques can effectively identify the most influential independent features on a variable of interest. This may help the modeller overcome the problem of degrees of freedom, which is one of the most encountered problems when modelling high dimensional data. It is observed that variable selection techniques have been used for two purposes: as an alternative to factor models such as LASSO (Least Absolute Shrinkage and Selection Operator) and LARS (Least-Angle Regression) to directly model the variable of interest (Konzen & Ziegelmann, 2016; Medeiros & Mendes, 2016) or as a pre-selection method in factor models (Bai & Ng, 2008; H. H. Kim & Swanson, 2018) to identify the targeted variables. Even though numerous papers have reported that ML models provide outstanding performance in the context of time series analysis (Ahmed et al., 2010; Pavlyshenko, 2019), the usage of these models is still limited in the field of macroeconomic forecasting (Masini et al., 2021; Medeiros et al., 2021).

The lack of interpretability of ML models and the difficulty of the problem of balancing the bias-variance trade-off may be regarded as two reasons for this. Making ML models more explainable and interpretable has recently become an active and very popular field of study and many different techniques have been proposed for this purpose (Carvalho et al., 2019; Linardatos et al., 2020). Among these techniques, Local Interpretable Model-agnostic Explanations (LIME) proposed by Ribeiro et al. (2016) Shapley Additive exPlanations (SHAP) by Lundberg and Lee (2017) have stood out and found many applications in diverse fields (ElShawi et al., 2020; Gramegna & Giudici, 2021; Mokhtari et al., 2019). With the help of these techniques, it is possible to gain insights into the predictions produced by ML methods by quantifying the contribution of each predictor in an additive manner. This enables us to understand which indicators or drivers most influence predictive performance.

The contributions of this study to the literature can be summarised as follows:

- First, previous studies have used a linear forecast equation and focused on factor extraction methods and variable selection to attain more accurate forecasts. However, the form of the forecast equation is an important though sometimes neglected component. In this study, the forecast equation is changed to a non-linear form with the help of ML models.
- Second, the usefulness of feature selection techniques in a high dimensional environment is investigated in two ways. First by examining the ability of different selection methods to determine a preferred set of targeted variables with predictive power for factor models. The second approach is to employ feature selection techniques as a pre-processing step in directly forecasting the variable of interest without using a factor model. This study reports an analysis and comparison of the two alternative approaches.
- The third main contribution is to take advantage of recent advancements in ML in order to explain the inflation predictions of these black-box models. This relates to a previous paper by Joseph et al. (2021) that deals with black boxes and examines the drivers of forecasts of UK inflation. Our study is the first to focus on explaining the drivers of inflation forecasts for an emerging economy. This involves identifying both the key informative variables but also the sign and the magnitude of their effects on each individual prediction. Our paper differs from that of Joseph et al. (2021) in two ways: by including additional ML models in the analysis and by proposing the integration of feature selection methods with ML models to address the problem of model interpretability in high-dimensional settings.
- The last novel contribution is to provide a more comprehensive analysis of tree-based models in the forecasting of inflation. This is based on recent finding by Medeiros et al. (2021) that indicate superiority of random forest models against other ML and benchmark models in literature. In addition, we analyse also Extremely randomized trees, Adaboost

trees, Gradient Boosting Decision Trees (GBDT) and eXtreme Gradient Boosting (XGBoost).

The rest of the paper is organised as follows. The next section is an overview of the literature on the application of factor and ML models to predict macroeconomic variables and also studies related to forecasting inflation in the Turkish economy. The subsequent section discusses the methodology to be followed, the details of data set and the models in the analysis. Section 4 presents the results in three parts in accordance with the aims of the study. Finally, Section 5 has concluding remarks.

2. Related Studies

This section is divided into two parts. The first part summarises the application of factor and ML models to macroeconomic forecasting. The second part provides information about previous studies of inflation forecasting especially in Turkey.

After the seminal studies carried out by Stock and Watson (2002a, 2002b) and Forni et al. (2000, 2005), interest in factor models has increased dramatically. Boivin and Ng (2006), and Bai and Ng (2008) claimed that using more variables to extract factors would not cause better forecasting performance and showed that the factors extracted from the smaller set of the variable pool by the help of a pre-selection method produced more accurate forecasts than the factors extracted from all variables. Banbura and Modugno (2014) utilised factor models in nowcasting and forecasting of the euro area gross domestic product (GDP). The datasets in the analysis consisted of three different compositions namely small (14 series), medium (48 series), and large (101 series) to investigate the effect of the size of the dataset on the performance of factor models. The obtained results showed that using small and medium dataset compositions led to more accurate results than the larger one.

In another study by Li and Chen (2014), the comparison between LASSO-based approaches and factor models was made in forecasting macroeconomic time series. Also, the combination of these two groups of models was analysed. The study found that LASSO-type models outperformed the factor models, and the combined forecasts gave promising results. Garcia et al. (2017) compared a factor model constructed by targeted predictors via a pre-selection procedure based on t-statistics in a linear model with shrinkage models, complete subset regression, and random forest in real-time inflation forecasting. They concluded that shrinkage models exhibited superior performances especially for short-horizon forecasts than other models considered. Medeiros et al. (2021) attempted to uncover the benefits of ML models in forecasting US inflation through a set of state-of-the-art modelling approaches in a data-rich environment. Their main finding is that the random forest model dominated all other models in the study. In a recent study by Joseph et al. (2021) forecasted UK inflation and tried to open the black box nature of ML methods by Shapley values and regression. Even though they utilised a limited number of non-linear ML models (Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), and random forest), it was suggested that ridge regression can be used by central banks and institutions forecasting inflation with large datasets. Ridge regression exhibited similar performance with ML models in long sample period but outperformed them in short sample period. It appears from the literature that shrinkage and some ML models were used both as an alternative to factor models in forecasting and a pre-selection method among all predictors for factor models (H. H. Kim & Swanson, 2018; Konzen & Ziegelmann, 2016; Li & Chen, 2014; Medeiros & Vasconcelos, 2016).

Regarding inflation forecasting in Turkey, Ogunc et al. (2013) carried out a wide range of econometric models from univariate models to multivariate models to find out which one is better at producing more accurate short-term inflation forecasts specifically for Turkey. The results reported suggest that models containing more information related to inflation through many variables outperform the univariate models in short-term forecasts. Altug and Cakmakli (2016) examined the power of survey expectations as a predictor in inflation forecasting for two emerging countries, Brazil and Turkey. The inclusion of survey expectations into models increased forecasting performances especially for Turkey which is more volatile compared to Brazil. Mandalinci (2017) analysed the predictability of inflation in nine emerging countries including Turkey by different modelling approaches and assessed the accuracy of institutional forecasts. It was found that central bank independence has a key role in the predictability of inflation and institutional forecasts are generally better than model-based ones for emerging markets. Soybilgen and Yazgan (2017) assessed inflation expectations in Turkey measured by a survey of experts and decision-makers to learn their predictions for current and future inflation. It was observed that inflation expectations do not have much predictive value and can be worse than even naïve methods, but the directional accuracy has some value.

Finally, Gunay (2018) made a comprehensive analysis of some significant components of factor models such as choice of feature extraction method, the number of factors and the number of lags of each factor in the forecast equation through datasets of industrial production and core inflation in Turkey. According to the study, the performance of factor models is dependent on the proper selection of important components. In a recent study by Ozgur and Akkoc (2021), shrinkage models (Ridge, Lasso, Adaptive Lasso, and Elastic net) were considered to have found the most effective predictors in predicting Turkish inflation. The authors claimed that the study is the first one that applies ML models for forecasting Turkish inflation, but they actually traditional used regression-based models rather than pure ML techniques like SVM or MLP. They found that the shrinkages models led to superior forecasts than autoregressive integrated moving average (ARIMA) and multivariate vector autoregression (VAR) models.

3. Models

3.1. Benchmark Models

The benchmark in this study is autoregressive models, AR(p), which use only the lagged values of the target variable

$$\hat{y}^{h}_{t+h/t} = \hat{\alpha}_h + \sum_{j=1}^{p} \hat{\gamma}_{hj} y_{t-j+1} \quad (1)$$

where y_t is the dependent variable, h represents the forecasting horizon, and p corresponds to the number of lags to be set a value of maximum three.

3.2. Feature Selection Techniques

Several shrinkage models are employed in this study for two purposes: for feature selection and as regressors for prediction. These models minimise the general penalised regression defined as follows:

$$\hat{\boldsymbol{\beta}}_h = arg \min_{\beta_h} \left[\sum_{t=1}^{T-h} (y_{t+h} - \boldsymbol{\beta}'_h \boldsymbol{x}_t)^2 + \sum_{i=1}^{M} p\left(\beta_{h,i}; \lambda\right) \right] \quad (2)$$

where T is the total number of observations, M is the number of predictors, $p\left(\beta_{h,i}; \lambda\right)$ corresponds to a penalty function whose tuning parameter is λ with the task of governing the amount of shrinkage.

Ridge Regression: Ridge regression, proposed by Hoerl and Kennard (1970), is one of the most well-known penalised regression models with L_2 norm penalty given by Equation 3. The intention behind ridge regression was to cope with highly correlated regressor by reducing the variance of the estimator in exchange for allowing a small bias. However, the coefficients obtained by ridge regression are very small but not exactly equal to zero. It means that it is not appropriate to select features. Hence, we exploit it as a prediction model but defined here to be compatible with other shrinkage models.

$$\sum_{i=1}^{M} p(\beta_{h,i}; \lambda) := \lambda \sum_{i=1}^{M} \beta_{h,i}^2 \tag{3}$$

LASSO: The LASSO estimator, introduced by Tibshirani (1996), offers sparse coefficients by using the penalty function called L_1 norm given in Equation 4, thereby overcoming the problem of ridge regression failing to equate the coefficients to zero. In this sense, LASSO can be regarded as a feature selection technique as well. It even works under the condition that the number of features is bigger than the total number of observations by finding a correct subset of the relevant features.

$$\sum_{i=1}^{M} p(\beta_{h,i}; \lambda) := \lambda \sum_{i=1}^{M} |\beta_{h,i}| \tag{4}$$

Elastic Net Regression: Elastic net, proposed by Zou and Hastie (2005), combines the L_1 and L_2 norms via Equation 5 to benefit from the useful properties of ridge and LASSO regression simultaneously. When there are many highly correlated features, LASSO has a tendency of picking just one randomly from this group and ridge regression can lead to better performances (Tibshirani, 1996). Elastic net embodies both models as special cases by using $\alpha \in [0,1]$ hyperparameter which controls the importance of L_1 and L_2 norms.

$$\sum_{i=1}^{M} p(\beta_{h,i}; \lambda) := \alpha\lambda \sum_{i=1}^{M} \beta_{h,i}^2 + (1-\alpha)\, \lambda \sum_{i=1}^{M} |\beta_{h,i}| \tag{5}$$

LARS: The LARS algorithm, proposed by Efron et al. (2004), is similar to forward stagewise regression in terms of selecting the most correlated variable with the residual of the previous regression model at each iteration. Unlike stagewise regression, in the case of a set of equally correlated features with the residual, it moves equiangularly between features in that iteration. There will be k ordered features after k iterations, and it includes LASSO as a special case. However, LARS algorithm is more prone to containing the effect of noise due to the procedure of refitting the residuals iteratively.

Orthogonal Matching Pursuit (OMP): the OMP algorithm (Pati et al., 1993) is an improved version of the matching pursuit algorithm (Mallat & Zhang, 1993) where its origin comes from the field of signal processing. The matching pursuit algorithm is devised to tackle the sparse optimisation problem as a greedy method. This algorithm selects one column (atom) from training dictionary atoms at a time to find out the best matching atom with the signal while minimising the approximation error for the signal iteratively. At each iteration, it finds the maximum correlation among the residuals from the previous approximation and the atoms to select the next atom. However, when the dictionary consists of similar atoms, the convergence will be very slow. OMP algorithm overcomes this situation by making the residuals orthogonal to the previously selected atoms with the help of Schmitt orthogonalization. OMP is computationally simple and efficient and emerges as an alternative to LASSO.

Recursive Feature Elimination with Random Forest (RFE_RF): Recursive feature elimination is a wrapper-type backward selection method for finding the subset of relevant features by taking model performance into consideration (Guyon et al., 2002). It is flexible to use any estimator provided that it has a property of ranking features via some importance values or coefficients. It is shown that random forest can work very well with recursive feature elimination (Granitto et al., 2006). RF provides importance scores for features, but this does not mean that it identifies the feature combinations that lead to better accuracy. Recursive feature elimination starts with all features and ranks them according to the scores provided by the estimator. The worst features are discarded, and the analysis is repeated for the remaining features until a stopping criterion is met. The number of features to be selected is a very crucial hyperparameter that is determined by cross-validation.

Boruta: the Boruta algorithm is a wrapper feature selection method based on random forests to find out all relevant features rather than just non-redundant ones by a statistical test (Kursa et al., 2010). The core idea behind the algorithm is to generate the shadow feature, which is basically randomised copy of the original one, in order to eliminate the correlation between output and that feature, and to compare the importance values associated with the original feature and its shadow version in the extended dataset. In a random forests model, the number of cases where a feature has a larger importance value than the maximum *Z* score among shadow features is recorded, and these values are later used to decide which features are important statistically.

3.3. Forecasting Models

SVM: the SVM uses kernel functions to map data into a high dimensional space where the transformed data can be modelled linearly. Even though it was originally proposed to solve classification problems, it was adapted to regression problems by an ε-insensitive loss function (Drucker et al., 1997) and performed successfully in time series forecasting applications (K. J. Kim, 2003; Thissen et al., 2003). SVM is appealing because of its ability to minimise empirical error while limiting the complexity of the fitted model, thereby preventing overfitting.

MLP: Although the emergence of neural networks model, which was developed to mimic the parallel working principle of the human brain, dates back to earlier times, it has become more popular and found many applications in various fields after the advent of the backpropagation algorithm (Rumelhart et al., 1986). It has an advantage of non-linear and flexible approximation ability to any continuous function with any specified level of precision thanks to hidden neurons. However, it may suffer from overfitting problem and requires many hyperparameters to be tuned carefully.

Random Forests (RF): RF consists of decision trees which is simple, fast, non-linear, and non-parametric machine learning approach for both regression and classification purposes. As an ensemble method, RF was purposed by Breiman (2001) to alleviate the overfitting problem that regression tree frequently faces and to lower the large variance of forecasts. Training data are randomly sampled with replacement to increase diversity between forecasts and a random subset of features is selected for each tree to make them more unrelated to each other. The final decision of RF model is generally found by taking the average of the individual trees for a regression problem or by using majority voting for classification.

Extremely randomized trees: Another tree-based model called extremely randomized trees was proposed by Geurts (2006) to reduce the problem of high variance of decision trees. This model benefits from the strategy of perturb-and-combine effectively in constructing decision trees after observing the high variance arising from cut-point choice is responsible for the poor performance of tree-based models. Two differences of this model from RF model are to select the cut-point completely randomly while splitting a node instead of the most discriminative one and to use all training data rather than a bootstrap sample. Thus, it is expected that the final trees exhibit less dependence on the particular learning sample.

Adaboost: Adaboost algorithm, by Freund and Schapire (1996), was developed to decrease the variance and bias of predictions of a weak learner at the same time. The algorithm aims to improve its predictions in a sequential manner by focusing on the bigger errors of the previous step. It firstly constructs a very short decision tree called decision stump after giving equal weights to all observations. The next step predicts the errors of the previous learner model and adaptively weights them according to the size of errors. The final prediction is the sum of all predictions made by weak learners. The original algorithm uses decision trees as base learners, but the idea can be applied to other models. This study exploits the variant of Adaboost algorithm proposed by Drucker (1997) for regression among other variants. One shortcoming of the algorithm is that it is heavily influenced by outliers.

GBDT: GBDT, proposed by Friedman (2001), introduces a paradigm shift, called functional gradient descent. It consists of an additive model built by a stage-wise fashion in which a decision tree regression model is added at a time and the previous learners remain unaltered. Any loss function can be used providing that it is differentiable, and this loss function is minimised by the gradient descent method when new learners are added to the algorithm. Different from the Adaboost algorithm, the decision trees in gradient boosting are bigger and, its parameters are optimised so as to decrease the loss function value. The final output is the sum of the predictions of all weak learners.

XGBoost: XGBoost is the improved version of the gradient boosting algorithm, developed by Chen and Guestrin (2016), relying on the regression trees called CART (Classification and Regression Trees) as a weak learner model. It has been a very popular machine learning method due to its success in many competitions such as Kaggle and DataCastle by offering superior performances in both generalisation and speed. Its superior performance to the gradient boosting algorithm stems from two important features. The first one is that the loss function consists of the second-order Taylor expansion, meaning that Newton boosting rather than gradient boosting, and includes the regularisation term for controlling the complexity of the trees to avoid the overfitting problem. The second is to enable parallel computing. In addition, XGBoost allows the feature subsampling to make the trees uncorrelated and to speed up the algorithm. The shortcoming of the algorithm is to contain a relatively high number of hyperparameters that have a great impact on performance.

3.4. Shapley values and regression

Shapley values come from cooperative game theory, in which the total payoff is shared by players in accordance with their contribution in a coalition (Shapley, 1953). Six decades later, a brilliant analogy was made between players in a game and variables in a supervised model (Štrumbelj & Kononenko, 2010). While players divide the payoff among themselves, a prediction is divided into parts each of which is generated by a separate input variable. This enables us to quantify the contribution of each variable to an individual model prediction by sampling from the marginal distribution. As a model agnostic method, SHapley Additive exPlanations (SHAP) can be calculated for any supervised black-box model, and they stand out from other interpretability approaches such as LIME (Ribeiro et al., 2016) and DeepLIFT (Shrikumar et al., 2017) by being the only method to satisfy three desirable properties (Lundberg et al., 2017) namely local accuracy, missingness, and consistency. That is the reason why the Shapley method is preferred in this study.

Let $\hat{y}_{t+h}$ represent a prediction made for horizon h at time t. This prediction can be decomposed into its components via the Shapley values as follows:

$$\hat{y}_{t+h} = \phi_{t,0}^{h} + \sum_{j=1}^{M} \phi_{t,j}^{h} \equiv \Phi_{t}^{h} \quad (6)$$

where M is the total number of variables and $\phi_{t,0}^{h}$ is the Shapley value for the j^{th} variable. The $j = 0$ component corresponds to the mean model prediction for the training set.

The Shapley value of a feature is obtained by the weighted average contribution of that feature to the prediction for all possible coalitions (the average marginal contribution):

$$\phi_{t,j}^{h} = \sum_{S \subseteq M \setminus j} \frac{|S|!(M-|S|-1)!}{M!} [f(S \cup \{j\}) - f(S)] \quad (7)$$

where S is the coalition and $f(S)$ corresponds to the prediction value of that coalition.

Even though Shapley additive values are good for finding out the drivers of a prediction, they are just descriptive and do not provide any statistical inference from which one can be sure that a predictor has statistically significant contribution to the model predictions. One remedy for this shortcoming is put forward by Joseph (2019) called Shapley regression. It is based on an auxiliary regression model constructed on the predictions and the corresponding Shapley values as follows:

$$y_{t+h} = \phi_{t,0}^{h} + \sum_{j=1}^{M} \phi_{t,j}^{h} \beta_{t,k}^{h} + \epsilon_t \equiv \Phi_t^h B_t^h + \epsilon_t \tag{8}$$

where $\widehat{\epsilon_t} \sim N(0, \sigma_\epsilon^2)$ and $\beta_{t,0}^{h} \equiv 1$. The null hypothesis for the coefficients $\beta_{t,k}^{h}$ is given below:

$$H_0^k(\Omega): \{ \beta_{t,k}^{h} \leq 0\} \tag{9}$$

It should be noted from eq. (9) that only positive coefficients in eq. (8) show significance. A possible negative coefficient means that the concerned model fails to learn from that predictor. This auxiliary regression measures the alignment between the Shapley components and the variable of interest. The ideal situation occurs when the vector $B_t^h \equiv 1$. If the null hypothesis H_0^k is not rejected then the predictor does not show statistically significant co-movement with the predictions.

Although the coefficients $\beta_{t,k}^{h}$ shed some light on the degree to which the predictor at investigation and the prediction are linked, they do not give information regarding the direction or magnitude of the components of the Shapley values. For this reason, we define a summary statistic called weighted contribution coefficient (WCC) given as below:

$$WCC_{X_k} = sign\left(r_{\phi_k X_k}\right) \frac{\langle|\phi_k(\hat{f})|\rangle_\Omega}{\sum_{k=1}^{M}\langle|\phi_k(\hat{f})|\rangle_\Omega} \quad \epsilon[-1,1] \tag{10}$$

where $r_{\phi_k X_k}$ is the correlation coefficient between the variable of interest X_k and its Shapley value ϕ_k and where $\langle . \rangle_{\Omega_k}$ represents for the average of x_k in $\Omega_k \epsilon R$.

This summary statistic shows the magnitude and direction of the average impact of each variable on the output. $\sum_{k=1}^{M}\left|WCC_{X_k}\right| = 1$. It is different from the Shapley share coefficients defined by Joseph (2019) because it does not depend on the $\beta_{t,k}$ coefficient in the linear regression model (3) since when the number of variables is high and/or the amount of observations is low, the $\beta_{t,k}$ coefficient is not reliable.

There are some pitfalls when using the Shapley values. The calculation of the Shapley value is very time-consuming due to the combinatorial calculation for all possible coalitions and, similar to other permutation-based methods, it may include unlikely instances because of random sampling from marginal distribution especially when there are some correlated features. For every feature there will be a corresponding Shapley value so, in this regard, it can be said that it is not a sparse explanation method. To overcome some of these pitfalls, Lundberg et al. (2018) put forwarded the TreeSHAP method which is much faster and takes the dependencies of features into consideration by relying on the conditional expectation instead of the marginal one. However, it works just for tree-based models and it estimates the Shapley value which is not equal to zero even for a feature that has no effect on the prediction as the original method does (Janzing et al., 2020; Sundararajan & Najmi, 2020).

We integrate different feature selection techniques with the calculation of Shapley values to avoid assigning Shapley values to features with no prediction power and to provide a sparse and simple explanation. Hence, at each estimation point, a feature selection technique is firstly applied to identify the variables with high prediction power on the predictor or to select a variable from a set of correlated variables, assigning a Shapley value of zero to the unselected variables. This is intended to reduce the possibility of including unrealistic covariates in the analysis and the risk of estimating non-zero Shapley values for uninformative variables.

4. Dataset and Pre-processing

It is reasonable to consider that inflation has a potential relationship with practically all economic indicators. This study exploits many candidate indicators from different contexts, such as interest rates, confidence indexes, stock exchanges, exchange rates, production, etc. In order to achieve a seamless transition to the policy of inflation targeting, Turkey abandoned the fixed exchange rate system after the 2001 economic crisis. Eventually, in 2006, Turkey officially announced that it started to implement the inflation targeting regime. This can be regarded as a real milestone for the whole economy which leads to structural transformations and changes the relationship between macroeconomic variables. Taking into account that some time series are not available for 2006, 1 January of 2007 was chosen to be the start of the sample for our analysis. The core inflation price index, which excludes energy, food and non-alcoholic beverages, alcoholic beverages, tobacco, and gold level, was chosen to be the target variable in line with the Central Bank of the Republic of Turkey‘s (CBRT) statement of establishing monetary policies according to core inflation. The core inflation is more stable compared to headline inflation and robust to uncontrollable price fluctuations and seasonal variations. The indicators used in the analysis and their web sources are given in Table 1. The number of variables amounts to 25 in total, and they reflect the effect of different indicators from many sources such as economic, financial, and international data. All variables consist of monthly observations from January 2007 and to August 2021.

Figure 1 presents monthly and yearly core inflation rates in Turkey. As can be seen from Figure 1, with the beginning of the year 2017, inflation in Turkish economy has faced with two-digit rates and has been subject to more volatility than previous years. The test data start from January 2017, consisting of 56 observations, and the expanding window with refitting in every next observation of the test data has been implemented. This means that all models in the analysis were refitted 56 times across the test data in order to adapt new information or structural changes on the indicators as quickly as possible. The test set is the period in which the annual inflation rates are mostly above 10 percent, and it includes sudden and rapid increases in foreign exchange rates, which is assumed to be responsible for the increase in the general level of prices, due to fluctuations in the economy and the adverse effects of Covid-19. This makes a real challenge for the study to see whether the findings of the interpretability method for ML models are compatible with general economic precepts.

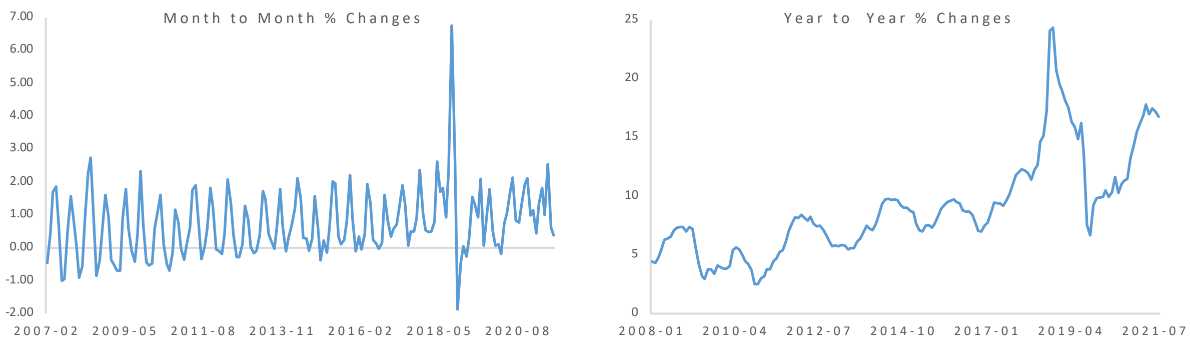

Figure 1. Core inflation rates.

Before the analysis, all series were checked for stationarity and seasonality. If a series exhibits seasonal components, the seasonally adjusted series was obtained by the X-13ARIMA-SEATS procedure, which is currently widely utilised by most government agencies all around the world, and the non-stationary series detected by the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test were made stationary by taking first or second differences. Lastly, the data were zero-centred and standardised before extracting factors.

5. Empirical Methodology

The advancement in technology especially for the last 50 years has led to the collection and storage of huge amounts of data in almost all areas of life. But it also introduces a new problem of how to extract valuable information from this high-dimensional data. In econometric modelling there will generally be a high number of highly correlated variables related to the variable to be forecasted. Due to the degree of freedom problem, ordinary least squares regression and vector autoregression models can contain a very limited number of independent variables and, hence, lack the ability to model complex relationships resulting from the interaction of many variables. Therefore, the development of methods that utilise large numbers of variables effectively is an active field of research.

Table 1. Variables used in the analysis.

Variables	Source
Economic Confidence Index	TURKSTAT
Consumer Confidence Index	TURKSTAT
Reel Sector Confidence Index	TURKSTAT
Industrial Production Index	TURKSTAT
Unemployment Rate	TURKSTAT
Core Inflation	CBRT
BIST-100 Index	CBRT
Capacity Utilisation Rate	CBRT
Commercial Credit Interest Rates for TL	CBRT
Commercial Credit Interest Rates for USD Dollar	CBRT
Consumer Credit Interest Rates for TL	CBRT
Euro/Dollar Parity	CBRT
Dollar Exchange Rate	CBRT
Interest Rates for Deposits in Turkish Lira	CBRT
Interest Rates for USD Dollar Deposits	CBRT
Bullion Gold Selling Price (TRY/Gr)	CBRT
Current Account	CBRT
Currency in circulation	CBRT
CPI Based Real Effective Exchange Rate	CBRT
12-months-ahead CPI survey of expectation (mean)	CBRT
Commodity Price Index	IMF
EU Consumer Confidence Indicator	EUROSTAT
EU Industrial Confidence Indicator	EUROSTAT
EU Production in Industry	EUROSTAT
S&P 500 Index	YAHOO-finance
VIX Index	YAHOO-finance

Note: Turkish Statistical Institute (TURKSTAT) web address is https://www.tuik.gov.tr/Home/Index. CBRT web address is https://www.tcmb.gov.tr/wps/wcm/connect/en/tcmb+en. International monetary fund (IMF) web address is: https://www.imf.org/en/Home. The statistical office of the European Union (EUROSTAT) is https://ec.europa.eu/eurostat/web/main/home. YAHOO-finance web address is https://uk.finance.yahoo.com/.

We address this with two different methodological approaches. Firstly, factor models were constructed by using all variables but four different forecast equations: linear, SVM, MLP, and RF. In this way, the classical form of the forecast equation in factor models was changed from linear one to non-linear form through the ML model considered in the study. Previous studies have ignored the form of the forecast equation and focused on data structuring and feature extraction techniques.

To select the targeted predictors from the pool of all candidate predictors, which will be used to generate the factors in the factor models, six feature selection techniques were implemented: LASSO, LARS, Elastic net regression (Elastic), OMP algorithm, RFE_RF, and also the Boruta algorithm. In total, this amounts to 28 factor models, most of which were relied on ML techniques, built in the analysis. Secondly, to take advantage of the ability of ML models in handling a high number of variables, and also to aid interpretability, ML models directly took the original variables without carrying out factor extraction. Again, the targeted predictors found by feature selection techniques were used to identify the most informative predictors. In this way, 21 ML models were built to model the macroeconomic variable. Lastly, instead of using shrinkage models as a feature selection approach, ML models were applied to forecast directly the core inflation by using their own equations. In addition to LASSO, LARS, Elastic, and OMP model, Ridge regression were used because of their promising performance reported in previous studies (Joseph et al., 2021; Li & Chen, 2014). The second part of this study extends the analysis of the tree-based models further relying on the recent findings (Medeiros et al., 2021) about the success of RF model in predicting key macroeconomic variables. Hence GBDT, Adaboost, extremely randomized trees, and XGBoost were employed.

It is highly likely that there is a lag relationship between the predictors and the variable to be forecasted in the analysis. To take this relationship into account, the lagged values of the predictors were taken as new predictors. The number of lagged values were decided on Bayesian Information Criteria (BIC) calculated for factor models. The study increased the lags of each predictor until four, and the minimum BIC value was obtained at three lags. Therefore, this resulted in 75 predictors, in total, for the analysis. It is known that the number of factors to be used in a forecast equation has a significant effect on the forecasting performance (Agostino & Giannone, 2012; Schumacher, 2007). Again, the BIC value was exploited to determine this value. It was observed that using only one factor had the smallest BIC value. Hence, the only one factor, the most explaining the variance, was chosen for all factor models constructed in the analysis. To tune hyperparameters of ML models and feature selection techniques, a modified 10-fold-cross-validation (the python package TimeSeriesSplit) for time series data in which the training sets that follow one another are supersets of the one that came before them, and the validation is always the last one in time was employed for the analysis. The models were estimated for inflation forecasts in a real-time environment but excluding data revisions to be made in the future. By real-time analysis, it is meant to utilise only the information available to the modeller at the time when the predictions are made. Multi-step-ahead forecasts are generated by taking the h-step-ahead predictive value to be the dependent variable, as suggested in the literature (Günay, 2018; Schumacher, 2007; Stock & Watson, 2002b).

The following is forecast equation for factor models:

$$\hat{y}^{h}_{t+h/t} = \hat{\alpha}_h + \sum_{j=1}^{p} \hat{\gamma}_{hj} z_{T-j} + \sum_{j=1}^{m} \hat{\beta}'_{hj} \hat{F}_{T-j+1} \tag{11}$$

$$z_t = 1200 \cdot (y_t - y_{t-1}) - 1200 \cdot (y_{t-1} - y_{t-2})$$

where y_t is the logarithm of the core inflation price index and $\hat{F}_T$ corresponds to the number of the first *k* estimated factors.

The numbers of p and m values were determined in terms of BIC value by limiting the maximum number of choices to four. To make h-step-ahead forecasts directly, the following transformation, by Stock and Watson (2002b), on the variable to be forecasted is defined:

$$y^{h}_{t+h} = \frac{1200}{h} \cdot (y_{t+h} - y_t) - 1200 \cdot (y_t - y_{t-1}) \tag{12}$$

To evaluate forecasting performances among different modelling approaches, we used relative versions of the root mean squared error (RMSE) and the mean absolute error (MAE), calculated on the test set. These relative calculations are based on the following equations, respectively:

$$Relative\ RMSE = \frac{RMSE(the\ relevant\ model)}{RMSE(AR(2))}, \qquad Relative\ MAE = \frac{MAE(the\ relevant\ model)}{MAE(AR(2))} \tag{13}$$

If these relative measures are less than 1, it indicates the forecasting performance of the relative model is better than the simple AR(2) model. For the sake of simplicity, we will call them the RMSE and MAE.

6. Empirical Results

6.1. The obtained results for factor, shrinkage, and ML models

This section presents the forecasting results of the benchmark models, various factor models arising from the combination of the form of forecast equation with feature selection techniques, shrinkage methods, and different ML modelling strategies for 1-, 3-, 6-, and 12-month-ahead horizons. In total, 58 forecasting models are evaluated. Table 2 lists the forecasting performances with respect to two error evaluation criteria: RMSE and MAE, respectively. Since these evaluation metrics vary considerably between the different forecasting problems, all evaluation criteria are referred to a benchmark which was taken to be the AR(2) model. To discern better performances than the benchmark more easily, values smaller than 1 are highlighted in bold.

In table 2 the set of statistically superior models with respect to each forecast horizon, which is detected by the model confidence set (MCS) test by Hansen et al. (2011), is represented by shadowing. The MCS tests for each horizon were done at 75% significance level, as suggested by Catania (2021). It is understood from this table that different error measures lead to change the set of superior models considerably for the horizons.

Among different factor model groups, each of which has different forecast equation form like linear, SVM, RF, and MLP, the RF factor model stands out from the other by producing more accurate forecasts than the benchmark for all evaluation metrics and forecast horizons and gives rise to more models to be included the set of the best-performing models. Furthermore, the other ML factor models, relying on the SVM and MLP, generally exhibit better performances than the linear factor models. This evidence supports the hypothesis that changing the functional form of a factor model may be beneficial. However, when feature selection techniques are applied to factor models, it is not clear how to determine the importance of individual input variables.

Regarding the performance investigation of shrinkage models, it can be said that the shrinkage models attain better performances than the benchmark model according to the RMSE possibly due to some extreme values observed in the price index. Hence, it may be concluded that the models relying on a high number of variables are less affected by the sudden increase or decrease in the dependent variable. Also, when we compare the shrinkage model with the classic linear factor models, it is seen that the shrinkage models are as good as or even better alternatives to the linear factor models. This observation is in line with the findings of some previous studies (Medeiros & Vasconcelos, 2016; Özgür & Akkoç, 2021). Another observation from Table 2 is that the MLP factor models generally are worse than the MLP models with a high number of variables. In other words, the MLP works very well in high-dimensional settings compared to the MLP factor models with a limited number of predictors. However, the same is not true for the RF models. They work effectively in both conditions, high and low input dimensionalities. Hence, there is no convincing evidence in these results that using factor models will produce more accurate forecasts. The effect of variable selection on model performance is also not clear but it should be noted that the RF with Lasso model is the only model which is always in the set of superior models for all forecast horizons and error measures. In general the RF model, which is less affected by feature selection technique, performs very well regardless of using summarised information provided by the factors or directly by modelling the original features.

Table 2. Forecasting performance for models with factors and independent covariates.

	RMSE				**MAE**			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
RW	1.259	1.219	1.066	1.014	1.520	1.289	1.157	1.130
AR(1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR(2)	**0.948**	1.069	1.017	1.044	1.109	1.217	1.077	1.102
AR(3)	**0.993**	1.176	1.041	1.031	1.151	1.287	1.086	1.098
Linear factor model by all variables	**0.969**	1.155	1.039	1.039	1.155	1.318	1.121	1.113
Linear factor model by LASSO	**0.960**	1.128	**0.995**	**0.867**	1.243	1.325	1.156	1.019
Linear factor model by LARS	**0.926**	**0.950**	**1.073**	**0.949**	1.224	1.105	1.171	1.066
Linear factor model by Elastic	1.019	1.170	1.008	**0.926**	1.374	1.323	1.108	1.085
Linear factor model by OMP	**0.898**	1.127	1.040	**0.887**	1.201	1.336	1.215	1.031
Linear factor model by RFE_RF	**0.970**	1.000	**0.928**	1.053	1.134	1.163	1.049	1.153
Linear factor model by Boruta	1.020	1.089	1.084	1.071	1.176	1.295	1.161	1.195
SVM factor model by all variables	**0.849**	**0.955**	1.016	**0.892**	**0.914**	1.085	1.075	**0.887**
SVM factor model by LASSO	**0.799**	1.100	**0.867**	**0.903**	**0.936**	1.048	**0.951**	**0.982**
SVM factor model by LARS	1.443	**0.809**	**0.912**	**0.852**	1.297	**0.917**	**0.969**	**0.864**
SVM factor model by Elastic	**0.821**	**0.900**	1.011	**0.918**	**0.954**	1.048	**0.994**	**0.932**
SVM factor model by OMP	1.149	**0.992**	**0.825**	**0.900**	1.119	1.041	**0.887**	**0.948**
SVM factor model by RFE_RF	**0.927**	**0.898**	**0.981**	**0.883**	**0.931**	1.033	**0.991**	**0.951**
SVM factor model by Boruta	1.139	1.548	1.227	**0.879**	1.044	1.395	1.180	**0.966**
RF factor model by all variables	**0.803**	**0.958**	**0.888**	**0.836**	**0.900**	**0.973**	**0.926**	**0.830**
RF factor model by LASSO	**0.846**	**0.964**	**0.952**	**0.832**	**0.973**	**0.975**	**0.972**	**0.818**
RF factor model by LARS	**0.814**	**0.963**	**0.885**	**0.831**	**0.935**	**0.966**	**0.900**	**0.818**
RF factor model by Elastic	**0.850**	**0.964**	**0.873**	**0.830**	**0.970**	**0.979**	**0.893**	**0.825**
RF factor model by OMP	**0.803**	**0.967**	**0.890**	**0.836**	**0.912**	**0.986**	**0.912**	**0.812**
RF factor model by RFE_RF	**0.819**	**0.957**	**0.893**	**0.833**	**0.914**	**0.962**	**0.910**	**0.824**
RF factor model by Boruta	**0.812**	**0.963**	**0.897**	**0.842**	**0.918**	**0.975**	**0.896**	**0.852**
MLP factor model by all variables	**0.854**	1.016	**0.956**	**0.906**	1.027	1.162	1.066	1.000
MLP factor model by LASSO	**0.884**	1.075	1.069	**0.875**	1.120	1.139	1.192	**0.961**
MLP factor model by LARS	**0.936**	1.075	**0.988**	0.914	1.159	1.118	1.032	**0.958**
MLP factor model by Elastic	**0.894**	**0.974**	**0.938**	**0.851**	1.131	1.081	1.073	**0.943**
MLP factor model by OMP	**0.908**	1.016	**0.896**	**0.941**	1.137	1.176	**0.964**	1.001
MLP factor model by RFE_RF	**0.820**	1.143	**0.984**	1.085	1.010	1.197	1.070	1.105
MLP factor model by Boruta	1.174	1.156	1.024	**0.807**	1.142	1.252	1.104	**0.896**
OMP	**0.884**	**0.987**	**0.941**	**0.893**	1.252	1.359	1.140	1.036
Lasso	**0.817**	1.021	**0.914**	**0.855**	1.023	1.200	1.118	**0.954**
Lars	**0.818**	1.025	**0.934**	**0.733**	1.029	1.216	1.055	**0.826**
Elastic	**0.783**	1.076	**0.929**	**0.861**	**0.966**	1.277	1.146	**0.964**
Ridge	**0.832**	**0.967**	**0.875**	**0.831**	**0.980**	1.031	**0.964**	**0.906**
SVM with all variables	**0.837**	**0.965**	**0.870**	**0.846**	1.009	**0.982**	**0.922**	**0.881**
SVM with LASSO	**0.913**	1.141	**0.816**	**0.869**	1.076	1.235	**0.926**	**0.939**
SVM with LARS	**0.891**	1.290	**0.847**	**1.003**	1.082	1.445	**0.991**	1.087
SVM with Elastic	1.212	1.029	**0.859**	1.095	1.387	1.143	**0.969**	1.065
SVM with OMP	**0.867**	1.227	1.016	1.213	1.029	1.214	1.157	1.088
SVM with RFE_RF	**0.783**	**0.946**	**0.837**	**0.865**	**0.926**	1.009	**0.945**	**0.906**
SVM with Boruta	**0.755**	**0.964**	**0.852**	**0.779**	**0.965**	1.021	**0.998**	**0.819**
RF with all variables	**0.824**	**0.966**	**0.900**	**0.808**	**0.948**	**0.958**	**0.946**	**0.831**
RF with LASSO	**0.816**	**0.974**	**0.866**	**0.829**	**0.915**	**0.965**	**0.909**	**0.840**
RF with LARS	**0.803**	**0.980**	**0.876**	**0.796**	**0.931**	**0.975**	**0.916**	**0.834**
RF with Elastic	**0.903**	**0.973**	**0.879**	**0.853**	1.002	**0.963**	**0.876**	**0.864**
RF with OMP	**0.796**	**0.968**	**0.879**	**0.841**	**0.929**	**0.949**	**0.955**	**0.850**
RF with RFE_RF	**0.822**	**0.954**	**0.901**	**0.811**	**0.948**	**0.951**	**0.951**	**0.828**
RF with Boruta	**0.816**	**0.972**	**0.884**	**0.843**	**0.929**	**0.958**	**0.919**	**0.853**
MLP with all variables	**0.831**	**0.979**	**0.822**	**0.864**	**0.979**	**0.958**	**0.998**	**0.905**
MLP with LASSO	**0.815**	**0.999**	**0.796**	**0.764**	**0.958**	1.084	**0.948**	**0.877**
MLP with LARS	**0.886**	**0.992**	**0.837**	**0.916**	1.081	1.036	**0.951**	**0.972**
MLP with Elastic	**0.867**	**0.964**	**0.824**	**0.834**	1.021	**0.987**	1.011	**0.868**
MLP with OMP	**0.851**	**0.973**	**0.881**	**0.993**	1.042	1.087	1.000	**0.982**
MLP with RFE_RF	**0.842**	**0.911**	**0.892**	**0.890**	1.030	1.030	1.000	**0.917**
MLP with Boruta	**0.905**	1.137	**0.810**	**0.848**	1.103	1.123	**0.969**	**0.943**

6.2. Comparison between tree-based models

The promising results of the RF models in the previous section and a similar finding by a recent paper (Medeiros et al., 2021) regarding inflation forecasting, led to the greater focus on tree-based models in the rest of this paper. In addition, it is observed that better or identical forecasts can be obtained through all or targeted features without using factor models. Moreover, even if we use a linear model constructed with latent factors by principal component analysis, it does not mean that its parameters have a clear economic interpretation (Buckmann et al., 2021). Since obtaining a more explainable prediction by machine learning is a central objective of this paper, we focus on tree-based models without factor modelling. Therefore, the aim of this section is to find out an answer to the question of whether it is possible to achieve more accurate forecasts by means of tree-based models constructed in various ways. For this purpose, in addition to the RF model, four tree-based models were used, and their interactions with feature selection techniques were analysed again. After adding the tree-based models to the analysis, in total, 86 forecasting models were constructed in the scope of this study.

Table 3 provides the forecasting performances of tree-based models for all horizons and error measures. We added the results of the RF model from the previous table to make comparison easy among tree-based models. Again, the MCS test was performed for each combination of horizon and error measure to identify statistically the set of the best-performing tree-based models shaded in the table. But here, the models better than the benchmark model are not denoted by bold fonts because almost all tree-based models with a few exceptions outperform the benchmark model. This is an indication of the ability of the tree-based models in dealing with big data for macroeconomic forecasting. As can be seen from Table 3, it may be concluded that the overall forecasting performance of tree-based models are very close to each other. It is hard to say that one method clearly dominates others in any specific horizon and error measure. But we notice that the performance of the extremely randomized trees is slightly worse than other ones according to the MCS test, especially for 1-month-ahead forecasts evaluated by both error measures and for 12-month-ahead horizon evaluated by MAE. Among the combination of tree-based models with feature selection techniques, some specific models such as RF with LASSO, Adaboost tree with OMP, and XGBoost with LARS are notable for leading to superior performance at all points of comparison. We will examine these models more closely in the next section to make explanations about their predictions.

Table 4 contains the average number of variables chosen by six feature selection techniques and their standard deviations for all horizons. As expected, the shrinkage models tend to select fewer variables. Among them, the OMP is the one that needs the minimum number of variables. The last two techniques in the table are tree-based selection methods. Especially, the RFE_RF is the most demanding one needing a large number of variables. It can be inferred that when the size of the forecasting horizon is increased from 1-month-ahead to 3-month-ahead and 6-month-ahead, there is a general tendency of selecting more variables. But for 12-month-ahead horizon, the number of selected variables decreases again for all techniques except for the RFE_RF.

It should be also noted that the model performances are closer to that of the benchmark model for 3- and 6-month-ahead horizons in which more features are employed by the competing models. The last two columns in Table 4 show the number of being in the set of superior models with feature selection techniques, which are calculated with respect to the results of Table 2 and 3. Note that the prevalence of well performing models with the LASSO is slightly higher than for other model selection methods, with respect to each error measure.

Table 3. The forecasting results of the tree-based models.

	RMSE				MAE			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
RF with all variables	0.824	0.966	0.900	0.808	0.948	0.958	0.946	0.831
RF with LASSO	0.816	0.974	0.866	0.829	0.915	0.965	0.909	0.840
RF with LARS	0.803	0.980	0.876	0.796	0.931	0.975	0.916	0.834
RF with Elastic	0.903	0.973	0.879	0.853	1.002	0.963	0.876	0.864
RF with OMP	0.796	0.968	0.879	0.841	0.929	0.949	0.955	0.850
RF with RFE_RF	0.822	0.954	0.901	0.811	0.948	0.951	0.951	0.828
RF with Boruta	0.816	0.972	0.884	0.843	0.929	0.958	0.919	0.853
GBDT with ALL	0.814	0.962	0.866	0.833	0.932	0.956	0.916	0.868
GBDT with LASSO	0.878	0.967	0.857	0.836	0.995	0.966	0.896	0.843
GBDT with LARS	0.822	0.993	0.879	0.826	0.964	0.992	0.920	0.825
GBDT with Elastic	0.883	0.970	0.851	0.826	0.998	0.958	0.896	0.799
GBDT with OMP	0.813	0.983	0.854	0.830	0.935	0.986	0.912	0.837
GBDT with RFE_RF	0.821	0.947	0.872	0.840	0.938	0.927	0.932	0.858
GBDT with Boruta	0.807	0.972	0.902	0.824	0.917	0.953	0.944	0.838
AdaboostTree with ALL	0.821	0.963	0.930	0.835	0.948	0.956	0.956	0.838
AdaboostTree with LASSO	0.841	0.976	0.861	0.705	0.960	0.951	0.896	0.807
AdaboostTree with LARS	0.844	0.978	0.931	0.826	0.984	0.976	0.951	0.812
AdaboostTree with Elastic	0.997	0.964	0.895	0.839	1.093	0.941	0.942	0.831
AdaboostTree with OMP	0.819	0.977	0.854	0.828	0.944	0.970	0.925	0.852
AdaboostTree with RFE_RF	0.821	0.975	0.925	1.016	0.949	0.944	0.978	0.972
AdaboostTree with Boruta	0.899	0.960	0.878	0.816	0.964	0.942	0.929	0.834
XGBoost with ALL	0.808	0.987	0.895	0.841	0.925	0.975	0.926	0.859
XGBoost with LASSO	0.801	0.973	0.854	0.827	0.910	0.962	0.900	0.835
XGBoost with LARS	0.804	0.977	0.880	0.828	0.937	0.963	0.918	0.847
XGBoost with Elastic	0.807	0.967	0.862	0.828	0.924	0.946	0.880	0.821
XGBoost with OMP	0.807	0.996	0.884	0.850	0.914	0.993	0.925	0.892
XGBoost with RFE_RF	0.807	1.002	0.913	0.852	0.935	0.951	0.913	0.856
XGBoost with Boruta	0.817	0.989	0.875	0.850	0.952	0.958	0.890	0.852
Ext.Rand.Trees with All	0.823	0.979	0.892	0.834	0.956	0.956	0.949	0.867
Ext.Rand.Trees with LASSO	0.831	0.990	0.883	0.873	0.965	0.976	0.900	0.883
Ext.Rand.Trees with LARS	0.838	0.978	0.890	0.851	0.965	0.979	0.935	0.877
Ext.Rand.Trees with Elastic	0.836	0.966	0.897	0.813	0.958	0.948	0.925	0.862
Ext.Rand.Trees with OMP	0.817	0.972	0.840	0.826	0.933	0.975	0.896	0.835
Ext.Rand.Trees with RFE_RF	0.827	1.011	0.887	0.866	0.959	1.013	0.958	0.903
Ext.Rand.Trees with Boruta	0.826	0.991	0.860	0.887	0.964	0.979	0.902	0.893

Table 4. The descriptive statistics for the variable selection techniques.

	h=1		**h=3**		**h=6**		**h=12**			
	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.	RMSE count	MAE count
LASSO	8.82	16.5	27.1	20.6	13.25	9.96	7.88	8.54	31	30
LARS	5.36	5.86	11.5	8.54	8.34	6.93	7.18	5.02	25	27
Elastic	6.16	12.7	27.7	19.0	15.8	10.3	9.91	9.60	28	29
OMP	2.84	1.88	3.80	2.38	3.39	2.58	1.27	0.73	26	25
RFE_RF	53.9	19.5	33.7	20.9	38.5	29.2	47.2	26.7	28	28
Boruta	11.1	9.89	7.04	3.65	8.25	4.38	6.29	4.77	27	28

6.3. Model-agnostic interpretability with Shapley values

After observing the success of the tree-based models for inflation forecasting with high-dimensional data, we now focus on the best models to explain their predictions locally and to gain insights about the general relationship between variables and predictors globally. For ease of reference, we gathered the models with statistically superior performance for all horizons and error measures in Table 5. In this table, there are 11 models and only one model of them without any feature selection, Adaboost tree with All, exhibits superior performance. This shows that the feature selection techniques help to obtain models that perform well in general. The RF_RFE, which is a greedy algorithm in feature selection, does not lead to any model to be included in this table. It is interesting that no model from the GBDT is able to produce superior performance in all situations. Among feature selection methods, LASSO and OMP are the most effective, each with three models included in the table.

Table 5. Best-performing tree-based models for all horizons and error measures.

	RMSE				MAE			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
RF with LASSO	0.816	0.974	0.866	0.829	0.915	0.965	0.909	0.840
RF with LARS	0.803	0.980	0.876	0.796	0.931	0.975	0.916	0.834
RF with OMP	0.796	0.968	0.879	0.841	0.929	0.949	0.955	0.850
RF with Boruta	0.816	0.972	0.884	0.843	0.929	0.958	0.919	0.853
AdaboostTree with ALL	0.821	0.963	0.930	0.835	0.948	0.956	0.956	0.838
AdaboostTree with LASSO	0.841	0.976	0.861	0.705	0.960	0.951	0.896	0.807
AdaboostTree with OMP	0.819	0.977	0.854	0.828	0.944	0.970	0.925	0.852
XGBoost with LASSO	0.801	0.973	0.854	0.827	0.910	0.962	0.900	0.835
XGBoost with LARS	0.804	0.977	0.880	0.828	0.937	0.963	0.918	0.847
XGBoost with Elastic	0.807	0.967	0.862	0.828	0.924	0.946	0.880	0.821
Ext.Rand.Trees with OMP	0.817	0.972	0.840	0.826	0.933	0.975	0.896	0.835

To select the models to be explained from Table 5, two criteria are applied. One of them is the average number of variables used by each model. As mentioned before, the expanding window scheme was implemented by refitting all models for every observation of test data. Thus, this average value indicates whether the concerned model is sparse or greedy one in dealing with the high-dimensional problem. A simple model that utilises fewer variables can provide more advantage in making interpretation by avoiding the noise caused by irrelevant or superfluous variables even if the performance of complex model is similar to the simple one. The other criterion is to have a small error value as much as possible. For this reason, the graphs in Figure 2 are drawn with axes showing the values of two criteria.

According to Pareto optimality, for h=1, RF with OMP and XGBoost with LASSO are in the Pareto front in terms of MAE. In other words, they are the non-dominated with respect to two criteria. Similarly, RF with OMP for h=1 in terms of RMSE is the only model dominated all others with respect to these criteria. From Figure 2, it appears that RF with OMP and XGBoost with Elastic for h=3, Ext.Rand.Trees with OMP and XGBoost with Elastic for h=6, and Adaboost tree with LASSO and Ext.Rand.Trees with OMP for h=12 are the non-dominated ones. We choose these models to open their black box by Shapley values.

Figure 3 indicates the relative importance values of each predictor obtained by WCC_{X_k} statistic for all horizons and the selected models according to the previous analysis. The x-axis of these figures is represented by the relative feature importance value vertically sorted and the y-axis shows the features whose importance value are greater than the threshold value of $\left|WCC_{X_k}\right| \geq 0.05$. WCC_{X_k} determines, in average, how much of model predictions can be explained by the predictor examined X_k and also shows what the direction of the relationship is between the predictor and predictions. The colours of the horizontal bars denote the direction: red bars illustrate positive relationship and blue bars describe negative relationship, respectively.

It can be seen from figure 3 that y(t-1) is the most influential feature for prediction in all ML models and horizons. This result can be expected because the own lags of a series has usually the most dominated impact on the overall behaviour of a series. However, in general, the direction of the contribution is negative for 1-month-ahead horizon, but it changes to positive one when the length of the evaluation horizon is bigger than 1-month. Apart from the lagged values of inflation, the most important predictor is dependent on the horizon and ML model examined.

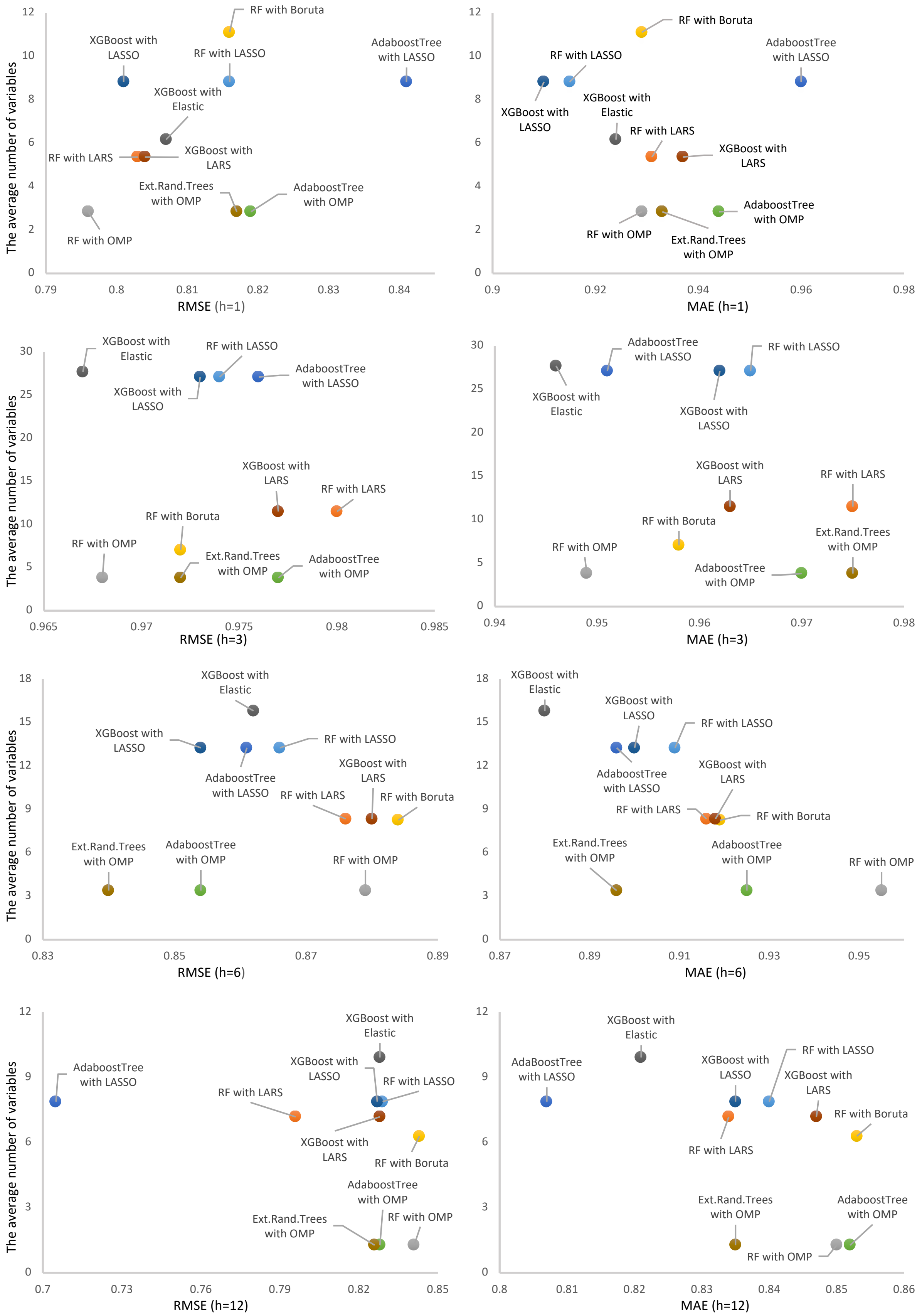

Figure 2. The scatter plot of two criteria for the best-performing tree-based models.

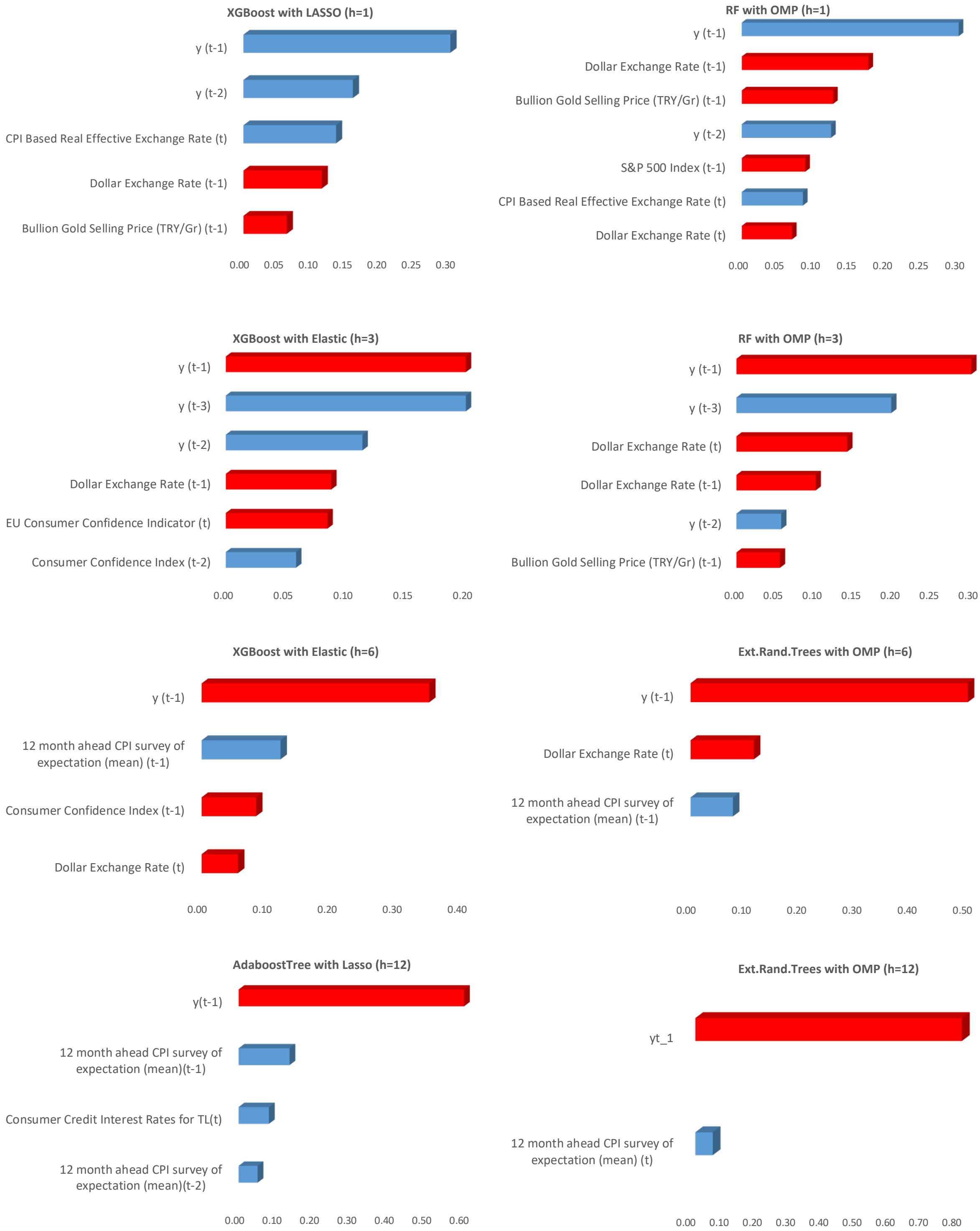

Figure 3. The variable importance bar plot for all horizons.

For short horizons (h=1 and h=3), the most important predictors in ML models are the Dollar exchange rate and a subset of variables mainly driven by foreign exchange rates such as bullion gold selling price, CPI based real effective exchange rate, and S&P 500 index. Considering the foreign currency dependency of the Turkish economy, this is in line with the general view on inflation. For longer horizons like h=6 and h=12, the best-performing ML models tell us that the effect of exchange rates is much reduced, and a smaller set of predictors has a key role in inflation forecasts. Finally, the direction of the relationship, detected by ML models via WCC_{X_k}, between predictors and inflation is consistent with the expectations.

After examining the general picture of the relative variable importance values and the direction of the associations between the predictors and inflation, we make a statistical inference by means of Shapley regression to identify which predictors have a statistically significant impact on the predictions of ML models. For this reason, the Shapley regression in Equation 14 is constructed in the scope of this study. This model restricts the number of predictors to be included by aggregating the predictors whose importance is low according to a threshold value as one predictor called remaining variables in the regression model. Thus, it will only focus on explaining the important variables.

$$y_{t+h} = \beta^h_{t,0}\phi^h_{t,base\_values} + \sum_{k=1}^{n}\beta^h_{t,k}\phi^h_{t,k} + \beta^h_{t,n+1}\phi^h_{t,remaining\_var} + \varepsilon_t \quad (14)$$

where n corresponds to the number of variables that satisfy the threshold value of $|WCC_{X_k}| \geq 0.05$.

Table 6 shows the results of Shapley regression for four horizons and the two best-performing ML models. This table presents the coefficients of the examined variables found by the linear model and the corresponding relative variable importance values given in parentheses. For each coefficient, $H_0^k(\Omega)\colon\{\beta^h_{t,k} \leq 0\}$ hypothesis is tested to find out the predictors' statistical significance in explaining inflation forecasts. The variables with a statistically significant contribution to the predictions of ML models are represented by bold font, and their significance levels are given at the bottom of the table. There are some negative coefficients meaning that the related variable is not learned well by the model in question. For 1-month-ahead horizon, only Dollar exchange rate (t-1) and the variable that is very closely related to it, CPI based real effective exchange rate (t), are found to be statistically significant predictors of ML models.

These findings are consistent with expectation. Since Turkey, as an emerging country, is heavily dependent on imported intermediate goods and meets most of its energy needs from abroad, the Turkish economy can be very sensitive to price fluctuations in Dollar exchange rate. Moreover, the signs of the WCC_{X_k} statistic given in parentheses for these variables are in harmony with the prevailing view in economics. As an indicator of the international trade competitiveness of a country, a decrease in real effective exchange rate increases the value of imported goods in domestic currency, thereby causing the consumer price index to be on the rise. Dollar exchange rate (t-1) has positive influence on the forecasts, and as a result of this, it leads to increases on the inflation expectation by ML models. In short horizons (h=1 and h=3), although more variables are included in the class of important variables, the number of significant variables is very low. Maybe high volatility and more noise in short horizons are responsible for this.

For all horizons longer than h=1, the lagged values of the series become significant predictors with a high predictive power in improving forecasting performance. In addition, it is seen that CPI survey of expectation turns out to have significant power on the models for 6- and 12-month-ahead horizons even if this power is inversely related to the inflation forecasts. A notable result is that consumer credit interest rates for TL (t) is found to be statistically significant predictor for h=12 by Adaboost tree with LASSO. For the data from Turkey, where the relationship between inflation and interest rates has been a topic of discussion recently, it is observed that interest rate hikes may have a reducing effect on inflation forecasts/expectations in the relatively long-term horizon (h=12). This can be an indication of demand inflation, showing an inverse relationship between inflation and interest rates in the long run.

The relationship between predictors and the forecasted variable may not hold over time especially when it comes to macroeconomic forecasting because some economic and non-economic events such as political and financial crises, recessions, and epidemic diseases like

COVID-19 can disrupt or even eliminate the structure that exist between variables. To investigate the overall changes in the variable importance over time, the graphs in Figure 4 are drawn for all horizon under examination. In these graphs, the test set is divided into six-month periods with one exception for the last period consisting of eight-month shown by the x-axis of the graphs. The y-axis in the graphs show the mean absolute Shapley values for the variables considered important according to the previous analysis. As can be seen from Figure 4, Dollar exchange rate and bullion gold selling price, which is highly positive correlated with exchange rates, come into play as substantially affecting predictors on the predictions of ML models specifically for 1-month-ahead forecasts after the period of January 2018 - June 2018. Although this effect is less, it still exists in other horizons except for h=12. During this period inflation and exchange rates surged, possibly linked to the decision to hold an early election in Turkey in the spring and foreign policy considerations. The dollar exchange rate rose to its historical highest value in the summer months and in 2018 and the Turkish lira depreciated by around 40% against the dollar. Following these events, the inflation rate reached approximately 25% annually in that year. Given these circumstances, it means that our forecasting framework quickly adapts itself to the new economic environment and successfully replaces its predictors with the new predictors which has more predictive power in that environment.

Table 6. Shapley regression for the best-performing models.

	h=1	
	XGBoost with LASSO	RF with OMP
y(t-1)	1.582 (-0.340)	-0.619 (-0.308)
y(t-2)	-0.712 (-0.159)	0.460 (-0.123)
Bullion Gold Selling Price (TRY/Gr) (t-1)	1.317 (0.063)	2.761 (0.126)
CPI Based Real Effective Exchange Rate (t)	**8.874** (-0.134)***	-1.433 (-0.085)
Dollar Exchange Rate (t)	-	6.828 (0.069)
Dollar Exchange Rate (t-1)	**9.054** (0.113)**	**5.236** (0.175)*
S&P 500 Index (t-1)	-	7.389 (0.088)
remaining variables	3.801 (0.192)	21.613 (0.025)
	h=3	
	XGBoost with Elastic	RF with OMP
y(t-1)	2.599 (0.235)	**4.396** (0.351)***
y(t-2)	0.173 (-0.114)	4.532 (-0.057)
y(t-3)	**3.368** (-0.232)*	1.631 (-0.198)
Bullion Gold Selling Price (TRY/Gr) (t-1)	-	1.408 (0.055)
Consumer Confidence Index (t-2)	1.769 (-0.059)	-
Dollar Exchange Rate (t)	-	5.120 (0.142)
Dollar Exchange Rate (t-1)	-2.788 (0.088)	-6.597 (0.101)
EU Consumer Confidence Indicator (t)	-3.439 (0.085)	-
remaining variables	1.116 (0.188)	-41.990 (0.096)
	h=6	
	XGBoost with Elastic	Ext.Rand.Trees with OMP
y(t-1)	1.193 (0.349)	**1.217** (0.586)**
Consumer Confidence Index (t-1)	3.392 (0.083)	-
Dollar Exchange Rate (t)	-1.112 (0.056)	0.080 (0.114)
12-month-ahead CPI survey of expectation (mean) (t-1)	1.945 (-0.121)	**11.351** (-0.076)***
remaining variables	**5.771** (0.391)***	-1.209 (0.224)
	h=12	
	AdaboostTree with LASSO	Ext.Rand.Trees with OMP
y(t-1)	**1.178** (0.606)***	**2.805** (0.856)***
12-month-ahead CPI survey of expectation (mean) (t)	-	-0.893 (-0.052)
12-month-ahead CPI survey of expectation (mean) (t-1)	**4.093** (-0.136)***	-
12-month-ahead CPI survey of expectation (mean) (t-2)	-9.404 (-0.051)	-
Consumer Credit Interest Rates for TL (t)	**6.389** (-0.081)***	-
remaining variables	0.982 (0.127)	-3.454 (0.048)

Note: *, **, and *** stand for the significance levels 0.1, 0.5, 0.01, respectively.

Note that the 12 month-ahead CPI survey of expectation (t-1) is significant for 6- and 12-month-ahead forecasts for extremely randomized trees with OMP and Adaboost tree with LASSO in Table 6. However, when we examine the corresponding graphs in Figure 4, it is found out that this variable has an impact on the predictions in only one period. It appears from Figure 4 that the lags of the core inflation play a more persistent dominant role as a predictor in longer horizons compared to short horizons. Lastly, even though consumer credit interest rates for TL (t) is the statistically significant predictor for h=12, we see that this variable only makes substantially contribution to the predictions in the dates between 07/2019 and 12/2019. The reason behind this that the CBRT decided to significantly increase interest rates in the period of 07/2018 and 12/2018, as a result of sharp increases in foreign exchange and inflation rates. Thus, our forecasting framework captures this phenomenon in producing more accurate and explainable forecasts.

In addition to global interpretation, the ability of the Shapley values to provide local explanations makes it attractive to open the black-boxes that are ML models. To illustrate this, we chose two samples from test set in which the XGBoost with LASSO is used to forecast 1-step-ahead values of the core inflation index for the dates of 07/2018 and 06/2020. 07/2018 is the first month after early election in a sense of reflecting the reaction to the result of the election.

The first graph in Figure 5 belonging to 2018/07 shows three predictors making contribution to the prediction value of $f(x)= -0.082$. CPI based real effective exchange rate make biggest and positive contribution represented by red colour and y(t-1) and Dollar exchange rate(t-1) are the predictors with negative contribution represented by blue colour. $E[f(x)]$=0.028 is the expected value (prediction) when all predictors in the model equal to zero. 06/2020 is the month where the inflation is on the rise probably related to the worldwide restrictions due to COVID-19. The graph at the bottom of Figure 5 indicates that our prediction is 0.103 for this month and the expected prediction is 0.068. Our prediction consists of three components one of which is positive one and the rest of which are negative ones. For this month, the lags of inflation are the dominating predictors.

Figure 4. Overall variable importance over time.

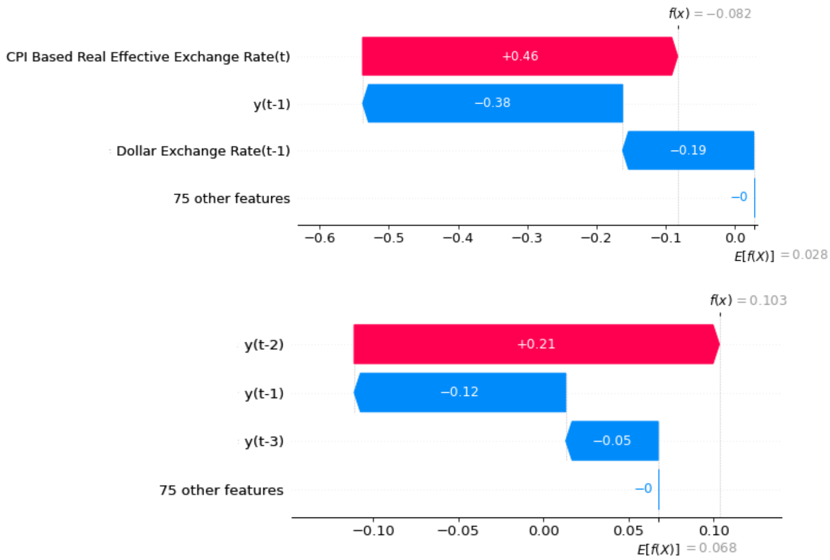

Figure 5. Local explanations by the XGBoost with LASSO on the dates of 07/2018 and 06/2020.

7. Conclusion

Considering the key role of short- and medium-term predictions for formulating monetary policies, it is of vital importance for central banks to produce their forecasts as accurately as possible. Traditionally, factor models have been employed in many macroeconomic problems. Recently, shrinkage models have become more popular among researchers when dealing with the problems involving many variables. However, these models are generally based on the assumption that the relationship between the response variable and predictors is linear. ML models offer an alternative modelling approach by taking nonlinear relationships into account.

Most studies in the literature exploit ML methods to attain better accuracy at the expense of the black box nature of ML models. However, central banks are responsible for clearly explaining the decisions they make. In this study, we utilise ML models both to achieve accurate forecasts and to provide explanations for individual predictions with high-dimensional inputs, by relying on Shapley values. For this purpose, a comprehensive comparison is made among different models in the data-rich environment of forecasting the core inflation index for Turkey. Some important conclusions can be drawn from our empirical findings, as follows.

First, our results show that using multivariable models provides an advantage in attaining more accurate forecasts than those of the univariate models. ML methods can be successfully implemented within the framework of the factor models by changing the form of the forecast equation from linear to nonlinear. These results are new in the inflation forecasting literature involving factor models.

Second, our results also show that ML methods can be directly applied in a macroeconomic forecasting problem in a high-dimensional setting by using all predictors, without the need for factor modelling. Among ML models, RF delivers regularly consistent forecasting performance in all empirical modelling methodologies in the study. But this is not true for the other ML models considered.

Third, an extended study of differently constructed tree-based models indicates that the general forecasting performance of tree-based models do not differ considerably except for the extremely randomized trees with slightly worse. In particular, the use of tree-based models with feature selection techniques delivers statistically superior set of forecasts for all horizons and error measures investigated. In this sense, it can be said that the obtained result contributes to the existing literature.

Finally, the integration of the tree-based forecasting framework of this study with Shapley values-based inference provides us a balance between improved model accuracy and statistical inference. For forecasting short horizons, it is seen that Dollar exchange rates and the predictors highly related with it are statistically identified to be significant in making contributions to the predictions of tree-based ML models. This conclusion is in line with expectations about the Turkish economy considering Turkey's economic structural characteristics.

Our analysis shows that it is possible to open the black box nature of ML models in an economic environment and to offer simple and sparse solutions to high-dimensional macroeconomic problems by addressing model interpretability. For future work, it is planned to compare with other state-of-the-art Explainable Artificial Intelligence (XAI) techniques to find which is optimal under which conditions. This will provide an opportunity for further interpretations about the predictions by measuring model fidelity. It will also better elucidate the general functional relationships between the covariates and inflation, to offer more evidence regarding the consistency of the results obtained with different modelling methodologies.

Funding: This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK).

References

Agostino, A. D., & Giannone, D. (2012). Comparing Alternative Predictors Based on Large-Panel Factor Models. *Oxford Bulletin of Economics and Statistics*, *74*(2), 306–326. https://doi.org/10.1111/J.1468-0084.2011.00642.X

Ahmed, N. K., Atiya, A. F., Gayar, N. El, & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Https://Doi.Org/10.1080/07474938.2010.481556*, *29*(5), 594–621. https://doi.org/10.1080/07474938.2010.481556

Altug, S., & Çakmakli, C. (2016). Forecasting inflation using survey expectations and target inflation: Evidence for Brazil and Turkey. *International Journal of Forecasting*, *32*(1), 138–153. https://doi.org/10.1016/J.IJFORECAST.2015.03.010

Alvarez, R., Camacho, M., & Perez-Quiros, G. (2012). Finite Sample Performance of Small Versus Large Scale Dynamic Factor Models. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2002663

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317. https://doi.org/10.1016/J.JECONOM.2008.08.010

Bańbura, M., & Modugno, M. (2014). MAXIMUM LIKELIHOOD ESTIMATION OF FACTOR MODELS ON DATASETS WITH ARBITRARY PATTERN OF MISSING DATA. *Journal of Applied Econometrics*, *29*(1), 133–160. https://doi.org/10.1002/JAE.2306

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*(1), 169–194. https://doi.org/10.1016/J.JECONOM.2005.01.027

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buckmann, M., Joseph, A., & Robertson, H. (2021). An interpretable machine learning workflow with an application to economic forecasting. *11th ECB Conference on Forecasting Techniques*.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics 2019, Vol. 8, Page 832*, *8*(8), 832. https://doi.org/10.3390/ELECTRONICS8080832

Catania, L. (2021). A Stochastic Volatility Model With a General Leverage Specification. *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2020.1855187

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672

Drucker, H. (1997). Improving Regressors using Boosting Techniques. In M. Kaufmann (Ed.), *In ICML Proceedings of Fourteenth International Conference on Machine Learning* (pp. 107–115).

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support Vector Regression Machines. In *Advances in Neural Information Processing Systems 9* (pp. 155–161). MIT Press.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*(2), 407–499. https://doi.org/10.1214/009053604000000067

ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2020). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*. https://doi.org/10.1111/COIN.12410

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics*, *82*(4), 540–554. https://doi.org/10.1162/003465300559037

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The Generalized Dynamic Factor Model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840. https://doi.org/10.1198/016214504000002050

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *In Proceedings of the 13th International Conference Om Machine Learning*, 148–156. https://ci.nii.ac.jp/naid/10024687385

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine on JSTOR. *The Annals of Statistics*, *29*(5), 1189–1232. https://www.jstor.org/stable/2699986

Garcia, M. G. P., Medeiros, M. C., & Vasconcelos, G. F. R. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, *33*(3), 679–693. https://doi.org/10.1016/J.IJFORECAST.2017.02.002

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/S10994-006-6226-1

Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. *Frontiers in Artificial Intelligence*, *4*, 140. https://doi.org/10.3389/FRAI.2021.752558

Granitto, P. M., Furlanello, C., Biasioli, F., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems*, *83*(2), 83–90. https://doi.org/10.1016/J.CHEMOLAB.2006.01.007

Günay, M. (2018). Forecasting industrial production and inflation in Turkey with factor models. *Central Bank Review*, *18*(4), 149–161. https://doi.org/10.1016/J.CBREV.2018.11.003

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning 2002 46:1*, *46*(1), 389–422. https://doi.org/10.1023/A:1012487302797

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The Model Confidence Set. *Econometrica*, *79*(2), 453–497. https://doi.org/10.3982/ecta5771

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Janzing, D., Minorics, L., & Blöbaum, P. (2020). Feature relevance quantification in explainable AI: A causal problem. *Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics*, 2907–2916. https://arxiv.org/abs/1910.13413

Joseph, A. (2019). *Parametric inference with universal function approximators* (No. 784; Staff Working Paper Series). https://www.bankofengland.co.uk/working-paper/2019/shapley-regressions-a-framework-for-statistical-inference-on-machine-learning-models

Joseph, A., Kalamara, E., Kapetanios, G., & Potjagailo, G. (2021). Forecasting UK inflation bottom up. In *Bank of England working papers*. Bank of England. https://ideas.repec.org/p/boe/boeewp/0915.html

Kim, H. H., & Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, *34*(2), 339–354. https://doi.org/10.1016/J.IJFORECAST.2016.02.012

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1–2), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Konzen, E., & Ziegelmann, F. A. (2016). LASSO-Type Penalties for Covariate Selection and Forecasting in Time Series. *Journal of Forecasting*, *35*(7), 592–612. https://doi.org/10.1002/FOR.2403

Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta – A System for Feature Selection. *Fundamenta Informaticae*, *101*(4), 271–285. https://doi.org/10.3233/FI-2010-288

Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, *30*(4), 996–1015. https://doi.org/10.1016/J.IJFORECAST.2014.03.016

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy 2021, Vol. 23, Page 18*, *23*(1), 18. https://doi.org/10.3390/E23010018

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777. https://doi.org/10.5555/3295222

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent Individualized Feature Attribution for Tree Ensembles. *ArXiv Preprint ArXiv:1802.03888*. http://github.com/slundberg/shap

Mallat, S. G., & Zhang, Z. (1993). Matching Pursuits With Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, *41*(12), 3397–3415. https://doi.org/10.1109/78.258082

Mandalinci, Z. (2017). Forecasting inflation in emerging markets: An evaluation of alternative models. *International Journal of Forecasting*, *33*(4), 1082–1104. https://doi.org/10.1016/J.IJFORECAST.2017.06.005

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2021). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 1–36. https://doi.org/10.1111/JOES.12429

Medeiros, M. C., & Mendes, E. F. (2016). $\ell 1$-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, *191*(1), 255–271. https://doi.org/10.1016/J.JECONOM.2015.10.011

Medeiros, M. C., & Vasconcelos, G. F. R. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, *138*, 50–52. https://doi.org/10.1016/J.ECONLET.2015.11.017

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business and Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745/SUPPL_FILE/UBES_A_1637745_SM2284.PDF

Mokhtari, K. El, Higdon, B. P., & Başar, A. (2019). Interpreting financial time series with SHAP values. *Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*, 166–172. https://dl.acm.org/doi/abs/10.5555/3370272.3370290

Öğünç, F., Akdoğan, K., Başer, S., Chadwick, M. G., Ertuğ, D., Hülagü, T., Kösem, S., Özmen, M. U., & Tekatli, N. (2013). Short-term inflation forecasting models for Turkey and a forecast combination analysis. *Economic Modelling*, *33*, 312–325. https://doi.org/10.1016/J.ECONMOD.2013.04.001

Özgür, Ö., & Akkoç, U. (2021). Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms. *International Journal of Emerging Markets*. https://doi.org/10.1108/IJOEM-05-2020-0577/FULL/XML

Pati, Y. C., Rezaiifar, R., & Krishnaprasad, P. S. (1993). Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 40–44. https://doi.org/10.1109/ACSSC.1993.342465

Pavlyshenko, B. (2019). Machine-Learning Models for Sales Time Series Forecasting. *Data*, *4*(1), 15. https://doi.org/10.3390/data4010015

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting*, *26*(4), 271–302. https://doi.org/10.1002/FOR.1026

Shapley, L. S. (1953). A Value for n-Person Games. In *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning Important Features Through Propagating Activation Differences. *In Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 3145–3153. http://proceedings.mlr.press/v70/shrikumar17a

Soybilgen, B., & Yazgan, E. (2017). An evaluation of inflation expectations in Turkey. *Central Bank Review*, *17*(1), 31–38. https://doi.org/10.1016/J.CBREV.2017.01.001

Stock, J. H., & Watson, M. W. (2002a). Macroeconomic Forecasting Using Diffusion Indexes. *Https://Doi.Org/10.1198/073500102317351921*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Stock, J. H., & Watson, M. W. (2002b). Forecasting Using Principal Components From a Large Number of Predictors. *Https://Doi.Org/10.1198/016214502388618960*, *97*(460), 1167–1179. https://doi.org/10.1198/016214502388618960

Štrumbelj, E., & Kononenko, I. (2010). An Efficient Explanation of Individual Classifications using Game Theory. *Journal of Machine Learning Research*, *11*, 1–18. http://www.ailab.si/orange/datasets.psp.

Sundararajan, M., & Najmi, A. (2020). The Many Shapley Values for Model Explanation. *Proceedings of the 37 Th International Conference on Machine Learning*, 9269–9278. https://arxiv.org/abs/1908.08474

Thissen, U., Van Brakel, R., De Weijer, A. P., Melssen, W. J., & Buydens, L. M. C. (2003). Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems*, *69*(1–2), 35–49. https://doi.org/10.1016/S0169-7439(03)00111-4

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/J.2517-6161.1996.TB02080.X

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/J.1467-9868.2005.00503.X