INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

Graduate Institute of International and Development Studies

International Economics Department

Working Paper Series

Working Paper No. HEIDWP25-2026

Forecasting Inflation in Tunisia Using Machine Learning Methods

Rawend Brahem
Central Bank of Tunisia

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

Forecasting Inflation in Tunisia Using Machine Learning Methods

Rawend Brahem
Central Bank of Tunisia

Abstract

Forecasting inflation in the presence of changing economic conditions, external shocks, and evolving transmission mechanisms remains an active area of research. This paper applies machine learning (ML) models to forecast headline and core inflation in Tunisia at 1-, 3-, and 6-month horizons, comparing their performance with standard econometric benchmarks within a rolling forecasting framework. Conformal prediction intervals are used to assess forecast uncertainty, while SHAP values serve as an exploratory tool to examine the contribution of explanatory variables to model predictions. The results reveal a clear horizon-dependent pattern, with the predictive gains of ML models increasing at medium and longer horizons. These gains are particularly pronounced at the 6-month horizon, where Support Vector Regression reduces the RMSE by more than 40% relative to the best-performing benchmark. Furthermore, SHAP analysis suggests that the relative contribution of predictors varies across forecasting horizons, with inflation persistence playing a more prominent role at short horizons and monetary, external, and commodity price variables becoming more relevant at longer horizons. Overall, these findings suggest that machine learning methods are most valuable as a complement to, rather than a substitute for, traditional forecasting approaches, particularly at longer horizons where nonlinearities become more pronounced.

Keywords: Inflation Forecasting, Machine Learning, Conformal Inference, SHAP Values, Tunisia

JEL: C53, E31, E37

The author thanks Professor Christophe Gaillac from the University of Geneva and Professor Marko Mlikota from The Geneva Graduate Institute for the academic supervision of this paper. This research took place through the visiting program under the Bilateral Assistance and Capacity Building for Central Banks (BCC), financed by SECO, and the Graduate Institute in Geneva.
The views expressed in this paper are solely those of the author and do not necessarily reflect those of the Central Bank of Tunisia.

1 Introduction

Accurate inflation forecasts play a central role in economic decision-making. For policymakers, inflation projections are a key input for the design and implementation of monetary and fiscal policies. For households, firms, and financial market participants, expectations about future inflation influence consumption, investment, wage negotiations, pricing strategies, and portfolio allocation decisions. Given the broad economic consequences of inflation fluctuations, improving forecast accuracy remains a fundamental objective in both macroeconomic research and policy analysis.

Despite its importance, inflation forecasting has long proven to be a challenging task. Faust and Wright (2013) compare the forecasting performance of various approaches and show that simple statistical benchmarks are often difficult to outperform in forecasting inflation. One explanation for this difficulty is that conventional models impose a fixed linear structure on relationships that are, in practice, time-varying and prone to structural breaks. When the underlying data-generating process shifts in response to unforeseen disruptions, models estimated on historical relationships become less reliable precisely when accurate forecasts are most needed.

These limitations have motivated growing interest in machine learning (ML) methods for inflation forecasting. Unlike conventional econometric models, ML techniques do not impose a fixed structural form and are specifically designed to handle high-dimensional datasets, nonlinear relationships, and complex interactions among explanatory variables, making them naturally suited to environments where traditional structural assumptions break down. A growing body of literature suggests that these methods can improve forecasting performance in macroeconomic applications. For example, Medeiros et al. (2021) show that ML models can outperform traditional benchmarks in forecasting the United States inflation using a large set of macroeconomic indicators. Similar evidence has emerged across several advanced and emerging economies, indicating that ML techniques can provide valuable tools for capturing the increasingly complex nature of inflation dynamics.

The potential advantages of ML are particularly relevant in environments characterized by structural change, evolving transmission mechanisms, and exposure to multiple sources of shocks. Tunisia provides a relevant case study in this context. Over the last two decades, the Tunisian economy has experienced periods of relative macroeconomic stability alongside episodes of domestic and external shocks. These developments have been accompanied by changes in inflation dynamics making inflation forecasting increasingly challenging.

Figure 1: Inflation Dynamics in Tunisia vs. the United States

Note: Monthly year-on-year (YoY) inflation rates for Tunisia and the United States over the sample period. The upper panel reports headline inflation, measured as the YoY percentage change in the Consumer Price Index (CPI), while the lower panel reports core inflation. Inflation rates are expressed in percentage points.

During the pre-2011 period, inflation remained relatively stable in Tunisia, generally fluctuating between 2 and 5 percent. Inflation cycles were short-lived, mean-reverting, and driven by temporary shocks. Headline and core inflation moved closely together, suggesting a close alignment between overall and underlying inflation dynamics during this period. Following the 2011 transition, inflation dynamics in Tunisia underwent structural shifts marked by altered persistence, heightened volatility, and a modified relationship between headline and core inflation measures. Exchange rate pressures and shifting macroeconomic conditions were linked to changes in inflation persistence and the duration of inflation episodes. Inflation cycles exhibited greater asymmetry and weaker mean reversion. Unlike advanced economies such as the United States, where inflation features lower persistence and faster adjustment toward long-run targets, Tunisia displayed more persistent inflation dynamics and heightened vulnerability to structural and external shocks.

These dynamics intensified during the 2020-2023 period, under the weight of the COVID-19 pandemic, global supply-chain disruptions, and severe commodity price fluctuations. Tunisia recorded its most pronounced inflationary episode during this period, driven largely by imported inflation, energy shocks, and surging food commodity prices. While the United States experienced a comparable inflation surge during this period, divergences emerged in the subsequent adjustment path. Inflation peaked earlier and declined more rapidly in the United States, whereas Tunisia experienced a delayed peak and a protracted disinflation process. Although price pressures moderated after 2024, the adjustment process has remained gradual.

Taken together, these developments suggest that inflation dynamics in Tunisia are characterized by structural changes, nonlinearities, regime dependence, and evolving transmission mechanisms, thereby underscoring the need to evaluate more flexible forecasting approaches. Against this background, this paper evaluates the ability of ML methods to forecast headline and core year-on-year (YoY) inflation in Tunisia at 1-, 3-, and 6-month horizons. The forecasting performance of several ML algorithms is compared with that of standard benchmark models. Beyond forecasting accuracy, the analysis considers two additional dimensions. First, forecast uncertainty is assessed through conformal prediction intervals. Second, the contribution of individual variables to ML forecasts is examined using SHAP (Shapley Additive Explanations), which decomposes individual predictions into variable-level contributions.

The paper contributes to the literature in several ways. First, it contributes to the body of research on ML-based inflation forecasting in Tunisia. Second, it extends the analysis beyond point forecasting by jointly examining predictive accuracy, forecast uncertainty, and exploratory model explainability.

The remainder of the paper is organized as follows. Section 2 provides an overview of the literature on the application of ML models to inflation forecasting, conformal inference and SHAP values. Section 3 presents the methodology and the data. Section 4 introduces the ML models and benchmark specifications. Section 5 reports the forecasting results and compares predictive performance across models and horizons. Section 6 examines the application of SHAP values in a time-series setting for inflation forecasting. Finally section 7 concludes the paper.

2 Literature review

The literature on inflation forecasting has evolved considerably over the past two decades, moving from traditional econometric approaches toward increasingly sophisticated machine learning (ML) techniques capable of handling large and high-dimensional datasets. Early contributions mainly relied on factor models and dimensionality reduction techniques, while more recent studies have explored a broad range of ML algorithms, including penalized regressions, tree-based models, neural networks, and ensemble methods.

One of the earliest strands of the literature emphasized the usefulness of large macroeconomic datasets for improving forecasting performance. Stock and Watson (2002) showed that factor models based on principal component analysis significantly improved the United States macroeconomic forecasts relative to univariate regressions, VAR models, and leading indicator approaches. Similar findings were reported by Artis et al. (2005) for the United Kingdom. These studies laid the foundations for subsequent ML applications by highlighting the importance of data-driven approaches and high-dimensional information sets in macroeconomic forecasting. With the growing availability of large datasets and increased computational capacity, ML methods progressively gained prominence in inflation forecasting. Early applications include Nakamura (2005), who applied neural networks to U.S. inflation forecasting, and Inoue and Kilian (2005), who explored penalized regression techniques such as Lasso and Ridge regression. Since then, the literature has expanded rapidly, particularly in advanced economies.

A growing body of empirical evidence suggests that ML methods often outperform traditional time-series models such as AR, VAR, and factor models. Chakraborty and Joseph (2017) showed that several ML algorithms improved inflation forecasts for the United Kingdom relative to conventional benchmarks. Medeiros et al. (2021) compared a wide range of forecasting approaches for U.S. inflation and found that Random Forests and Lasso-based models consistently outperformed standard econometric specifications. Similar conclusions were obtained for Japan by Maehashi and Shintani (2020), and more recently by Naghi et al. (2024) for the United States, the United Kingdom, and Canada, although the latter study highlights that ML performance may weaken during periods of heightened instability such as the post-2020 inflationary episode.

Although the literature initially focused on advanced economies, an increasing number of studies have extended ML techniques to emerging and developing countries. Evidence of improved forecasting performance has been documented in several contexts: García et al. (2017) for Brazil, Baybuza (2018) for Russia, Rodríguez-Vargas (2020) for Costa Rica, Shah et al. (2022) for Pakistan, Das and Das (2024) for India, and Araujo and Gaglianone (2023) for Brazil.

Overall, these studies confirm the growing relevance of ML techniques for inflation forecasting in emerging economies, although results remain sensitive to forecast horizons, structural breaks, and data availability. Despite this expanding literature, empirical applications remain heavily concentrated in advanced economies, particularly the United States, the United Kingdom, Japan, and the euro area. By contrast, fewer studies have considered developing-country contexts, including Tunisia, in the application of ML methods to inflation forecasting.

More recently, the literature has progressively moved beyond forecast accuracy alone toward issues related to forecast uncertainty and model interpretability. In this context, conformal prediction methods have attracted growing attention as flexible tools for constructing statistically valid prediction intervals without relying on strong distributional assumptions. Originally introduced by Vovk, Gammerman, and Shafer (2005), conformal inference was later extended by Lei et al. (2018), who proposed computationally efficient procedures for regression settings, while Barber et al. (2023) adapted the framework to more complex and dependent data structures such as time series. Angelopoulos and Bates (2023) provide a comprehensive overview of recent methodological developments and applications.

At the same time, the increasing use of complex ML algorithms has intensified concerns regarding model interpretability. Several studies have therefore focused on explainable artificial intelligence (XAI) techniques aimed at identifying the contribution of predictors within black-box models. Among the most widely used approaches are LIME by Ribeiro et al. (2016) and SHAP values introduced by Lundberg and Lee (2017). SHAP values originate from cooperative game theory and are based on the Shapley value framework developed by Shapley (1953), where the total payoff of a game is distributed among players according to their individual contributions. This framework was later adapted to ML by Štrumbel and Kononenko (2010), allowing model predictions to be decomposed into additive contributions associated with each explanatory variable. In economics and finance, SHAP values have become increasingly popular because they improve the explainability of ML models and help identify the main drivers underlying model predictions.

In the context of inflation forecasting, SHAP-based methods provide useful insights into the macroeconomic factors contributing to ML predictions and can enhance the interpretability of forecasting models. However, the literature has also highlighted important limitations related to feature dependence, especially in macroeconomic datasets where explanatory variables are often highly correlated. To better accommodate feature dependence, Lundberg et al. (2019) proposed TreeSHAP for tree-based models, while Aas, Jullum, and Løland (2021) extended the Kernel SHAP framework to better account for dependent features and grouped variable contributions.

Overall, the recent literature reflects a broader evolution in ML applications to macroeconomic forecasting, moving from a sole focus on predictive performance toward more reliable and interpretable frameworks. However, fewer empirical evidence combine ML forecasting, uncertainty quantification through conformal inference, and SHAP-based forecast decomposition in contexts such as Tunisia. The present study contributes to this literature by evaluating the performance of ML methods for inflation forecasting in Tunisia while integrating conformal prediction intervals and SHAP-based decomposition methods to better assess forecast uncertainty and the contribution of individual variables to model predictions.

3 Methodology

Let inflation dynamics be represented by the following forecasting framework:

$$\pi_{t+h} = F_h(x_t) + u_{t+h}, \quad h = 1, \ldots, H,\ t = 1, \ldots, T,$$

where π_{t+h} denotes either headline inflation or core inflation (excluding fresh food and administered prices) at period $t + h$. The vector $x_t = (x_{1t}, \ldots, x_{nt})'$ contains a set of explanatory variables based on the latest observations available at the forecast date t, including lagged inflation terms and a broad range of macroeconomic and financial indicators. The function $F_h(\cdot)$ represents the underlying relationship between the predictors and future inflation, while u_{t+h} denotes a zero-mean disturbance term. A separate forecasting function is estimated for each forecast horizon h.

To reduce dimensionality and identify the most informative lag structure, lagged predictors are selected using the Sparse Group Lasso approach proposed by Simon et al. (2013). This method extends the traditional Lasso framework by simultaneously performing variable selection at both the group and individual levels, making it particularly suitable for high-dimensional settings. In the present framework, each explanatory variable constitutes a group composed of its twelve monthly lags.

The analysis considers three forecasting horizons corresponding to short-, medium-, and longer-term inflation dynamics, namely $h = 1$, 3, and 6 months ahead. The predictor set includes nineteen numerical variables capturing different transmission channels of inflation, including indicators related to real economic activity and domestic demand, external and exchange rate conditions, external spillovers, and commodity price developments.

Inflation forecasts are generated using two different forecasting strategies depending on the class of models considered. For the ML models, a direct forecasting approach is adopted, whereby a separate model is estimated for each forecast horizon without forecasting the explanatory variables. By contrast, benchmark econometric models, namely the AR(1) and VAR(1) specifications, are implemented using a recursive forecasting approach in which multi-step-ahead forecasts are generated iteratively from one-step-ahead predictions.

The general forecasting equation can be expressed as:

$$\hat{\pi}_{t+h|t} = \hat{F}_{h,\, t-R_h+1:t}(x_t),$$

where $\hat{\pi}_{t+h|t}$ denotes the forecast of π_{t+h} formed at time t, $\hat{F}_{h,\, t-R_h+1:t}(x_t)$ denotes the estimated forecasting function obtained using observations from period $t - R_h + 1$ to t, and R_h represents the estimation window length associated with horizon h.

Model estimation and forecast evaluation are conducted using monthly data within a rolling out-of-sample forecasting framework. As new observations become available, the estimation sample is recursively updated while the evaluation window moves forward through time. This approach allows the models to adapt to evolving economic conditions and mitigates the impact of potential structural breaks and regime shifts.

The initial estimation window spans the period from the first available observation to December 2019, corresponding to 215 observations for headline inflation and 204 for core inflation. The estimation window is then shifted forward by one observation at each iteration, while maintaining a constant window size throughout the forecasting exercise.

Hyperparameter selection for the ML models is performed using time-series cross-validation procedures specifically designed to preserve the chronological ordering of the observations. Unlike conventional K-fold cross-validation, which assumes independently and identically distributed observations, time-series data exhibit strong temporal dependence that must be preserved during model validation. Accordingly, a time-split K-fold cross-validation approach is implemented, where the first k folds are used for training and the subsequent fold is reserved for validation. Each successive training sample therefore moves forward over time while maintaining a fixed window of previous observations. For every validation fold, the Root Mean Squared Error (RMSE) is computed, and the set of hyperparameters minimizing the average RMSE across folds is retained.

Hyperparameter tuning and Sparse Group Lasso-based variable selection are performed separately for each inflation measure and forecast horizon. This design allows both the set of relevant predictors and the degree of model regularization to adapt to horizon-specific and series-specific environments, thereby acknowledging potential heterogeneity in inflation dynamics across forecasting horizons.

The hyperparameters are optimized for each model class using the initial training sample. Starting in January 2020, the models are recursively re-estimated over rolling training windows using the selected hyperparameters, and out-of-sample forecasts are generated at horizons of 1, 3, and 6 months ahead for both headline and core inflation. Forecast accuracy is subsequently evaluated using the RMSE criterion.

The dataset used in this study covers the period from February 2002 to December 2025 for headline inflation, and from January 2003 to December 2025 for core inflation.

Data

The data consist of a monthly macroeconomic and financial variables designed to capture the main determinants of inflation dynamics in Tunisia. The dataset encompasses indicators related to monetary and financial conditions, external sector developments, real economic activity, commodity prices, and international spillovers. Variables originally available at a quarterly frequency were converted to monthly frequency using the Denton method. The dataset reflects the information set available at the time of the analysis and spans approximately 287 monthly observations for headline inflation and 276 observations for core inflation and brings together indicators that jointly reflect both domestic structural characteristics and external sources of inflationary pressure.

Table 1: Data Description

Variable	**Category**
Policy Rate	Monetary
Net Claims on Central Government	Monetary
Claims on the Economy	Monetary
Tunindex	Financial
EUR/TND	External
USD/TND	External
Foreign Assets	External
Trade Balance	External
Real GDP Growth	Real activity
Industrial Sale Price Indices	Real activity
Salaries evolution rates in the non-agricultural private sector	Real activity
Recruitment (Placements réalisés), in thousands	Real activity
Entries of Non-Residents	Real activity
Tourist Overnight Stays	Real activity
Real Estate Price Index	Real activity
Energy Index	Commodities
Wheat Prices	Commodities
Soybean Meal Prices	Commodities
Euro Area Inflation	External spillovers

Note: The data are drawn from three main sources: the Central Bank of Tunisia, the National Institute of Statistics (INS), and the World Bank.

4 Models

This section presents the forecasting models considered in this study. The benchmark models include a Random Walk model, an autoregressive (AR) model, and two vector autoregressive (VAR) specifications. The machine learning models include regularized linear models and nonlinear algorithms designed to capture potentially complex relationships between inflation and its determinants.

4.1 Benchmark Models

4.1.1 Random Walk

The Random Walk (RW) model assumes that the best forecast of future inflation is its most recently observed value. The forecasting equation is

$$\hat{\pi}_{t+h|t} = \pi_t,$$

where $\hat{\pi}_{t+h|t}$ denotes the forecast at horizon h based on information available at time t.

4.1.2 Autoregressive Model

The autoregressive model of order one [AR(1)] captures inflation persistence by relating current inflation to its first lag. The model is specified as

$$\pi_t = \phi_0 + \phi_1 \pi_{t-1} + u_t,$$

where ϕ_1 measures inflation persistence. Multi-step-ahead forecasts are generated recursively.

4.1.3 Vector Autoregressive Models

Vector autoregressive models (VAR) extend the autoregressive framework by allowing several variables to evolve jointly. The general VAR(1) specification is

$$y_t = A_0 + A_1 y_{t-1} + \varepsilon_t,$$

where y_t is a vector of endogenous variables, A_1 is the matrix of lag coefficients, and ε_t is a vector of disturbances.

Two VAR(1) specifications are considered. The first includes inflation, Real GDP Growth, and the policy rate, capturing interactions between inflation dynamics, economic activity, and monetary policy. The second additionally incorporates the EUR/TND exchange rate and energy index to account for external inflationary pressures and imported cost shocks.

4.2 Machine Learning Models

The models considered in this study are divided into two groups: regularized linear models and nonlinear ML methods.

4.2.1 Regularized Linear Models

Regularized regression estimate the linear relationship between the target variable and the predictors by adding a penalty term to the least-squares objective function. The penalty term controls the magnitude of the regression coefficients, with different penalty specifications leading to different regularization approaches. Three regularized models are considered:

- **Lasso Regression** (Tibshirani, 1996), which applies an L_1 penalty and can shrink some coefficients exactly to zero, performing variable selection.
- **Ridge Regression** (Hoerl and Kennard, 1970), which applies an L_2 penalty and shrinks coefficients continuously toward zero without setting them exactly to zero, improving estimation stability in the presence of correlated predictors.
- **Elastic Net Regression** (Zou and Hastie, 2005), which combines both L_1 and L_2 penalties, balancing sparsity and coefficient stability.

4.2.2 Nonlinear Machine Learning Models

The nonlinear models allow the forecasting function to capture complex interactions among economic variables.

- **Support Vector Regression (SVR)**, introduced by Vapnik (1995), estimates a flexible relationship between predictors and inflation using kernel functions. By mapping predictors into a higher-dimensional space, SVR can capture nonlinear relationships while controlling model complexity.
- **Gaussian Process Regression (GPR)**, developed by Rasmussen and Williams (2005), is a non-parametric Bayesian approach that models the forecasting function as a stochastic process. It provides both point forecasts and measures of predictive uncertainty.
- **Random Forest** (Breiman, 2001) is an ensemble learning method that combines multiple decision trees estimated from bootstrap samples and random subsets of predictors. Averaging across trees reduces variance and improves predictive robustness.
- **Extreme Gradient Boosting (XGBoost)**, introduced by Chen and Guestrin (2016) and based on the gradient boosting framework of Friedman (2001), is a sequential ensemble learning method where each new tree improves upon previous prediction errors. This structure allows the model to capture nonlinear relationships while maintaining strong predictive performance.

4.3 Feature Selection: Sparse Group Lasso

Given the large number of candidate predictors and lagged variables considered in this study, a feature selection procedure is implemented prior to model estimation in order to reduce dimensionality and mitigate potential overfitting. Specifically, the Sparse Group Lasso (SGL) is employed to identify the most informative predictors for inflation forecasting.

The SGL extends the Lasso and Group Lasso approaches by combining individual-level and group-level regularization. This feature is particularly useful in macroeconomic forecasting applications where predictors exhibit a natural grouped structure. In the present framework, each explanatory variable constitutes a group containing its twelve monthly lags.

Following Simon et al. (2013), the coefficient vector is obtained by solving:

$$\min_{\beta} \left\{ \frac{1}{2n} \left\| y - \sum_{l=1}^{m} X^{(l)} \beta^{(l)} \right\|_2^2 + (1-\alpha)\lambda \sum_{l=1}^{m} \sqrt{p_l} \left\| \beta^{(l)} \right\|_2 + \alpha\lambda \left\| \beta \right\|_1 \right\},$$

where $X^{(l)}$ and $\beta^{(l)}$ denote, respectively, the submatrix of predictors and the coefficient vector for group l, and p_l is the size of that group. The parameter $\lambda \geq 0$ controls the overall degree of regularization, and $\alpha \in [0, 1]$ balances the two penalties: $\alpha = 0$ gives the Group Lasso, and $\alpha = 1$ gives the standard Lasso.

Unlike the standard Lasso, which treats all coefficients independently, the Sparse Group Lasso explicitly exploits the grouped structure of the predictors. As a result, the method is capable of removing entire groups of variables while simultaneously selecting only the most informative lags within the retained groups.

5 Empirical results

5.1 Forecasting results

This section presents the forecasting performance of ML models relative to standard econometric benchmarks across 1-, 3-, and 6-month horizons for both headline and core inflation. The analysis reveals several robust patterns regarding horizon dependence, model performance, and the relative importance of nonlinearities.

Forecast accuracy is evaluated using the Root Mean Squared Error (RMSE), defined as:

$$RMSE_h = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\pi_{t+h} - \hat{\pi}_{t+h|t} \right)^2},$$

A primary finding, as reported in Tables 2 and 3, is that forecast errors generally increase with the forecasting horizon, reflecting the growing uncertainty associated with medium-term inflation dynamics. However, both the magnitude of this deterioration and the relative ranking of models differ substantially across inflation measures and model classes.

Model	1-month	3-month	6-month
Lasso	**0.273**	**0.663**	0.912
Ridge	0.331	0.688	0.913
Elastic Net	0.279	0.670	0.913
SVR	0.305	0.686	**0.672**
GPR	0.386	0.685	0.739
Random Forest	0.644	0.874	0.759
XGBoost	0.530	0.948	0.780
AR(1)	0.260	**0.612**	**1.130**
Random Walk	0.277	0.634	1.138
VAR (3 vars)	**0.238**	0.616	1.158
VAR (5 vars)	0.240	0.620	1.144

Table 2: Forecasting performance (RMSE) for Headline Inflation.

Model	1-month	3-month	6-month
Lasso	**0.194**	0.397	0.455
Ridge	0.198	0.395	0.487
Elastic Net	0.196	0.396	0.470
SVR	0.201	**0.358**	**0.415**
GPR	0.267	0.394	0.430
Random Forest	0.304	0.561	0.578
XGBoost	0.375	0.519	0.611
AR(1)	0.220	0.554	1.045
Random Walk	0.226	0.556	**1.028**
VAR (3 vars)	0.202	0.544	1.042
VAR (5 vars)	**0.200**	**0.539**	1.030

Table 3: Forecasting performance (RMSE) for Core Inflation.

Note: RMSE measures out-of-sample forecast errors at 1-, 3-, and 6-month horizons. Lower values indicate better forecasting performance. Bold values indicate the lowest RMSE within each forecasting horizon for ML and benchmark models.

At the 1-month horizon, traditional econometric models remain highly competitive, particularly for headline inflation. The VAR(1) model achieves the best performance (with an RMSE of 0.238), outperforming all ML approaches, the best of which (Lasso) records an RMSE of 0.273. In contrast, ML models provide more accurate forecasts for core inflation at the same horizon. Lasso achieves the lowest RMSE (0.194), closely followed by Elastic Net (0.196) and Ridge (0.198), all outperforming the best benchmark specification, VAR(1) (RMSE = 0.200), SVR performs comparably (0.201), while GPR displays a noticeably higher error (0.267) at this horizon. Overall, these results suggest that short-term headline inflation dynamics are largely captured by persistence and linear structures, which are effectively modeled by traditional econometric approaches, whereas core inflation forecasts benefit from the additional flexibility offered by ML methods, with regularized linear models performing particularly well at this horizon.

At the 3-month horizon, the best-performing ML model for headline inflation remains Lasso (RMSE = 0.663), while the best-performing benchmark specification is the AR(1) model (RMSE = 0.612). For core inflation, ML models continue to achieve lower forecast errors than traditional benchmark specifications, although the ranking among ML models shifts: SVR now delivers the lowest RMSE (0.358), followed by GPR (0.394) and the rest of the linear regularized models which display comparable, though somewhat higher, forecast errors.

At the 6-month horizon, ML models substantially outperform benchmark approaches for both headline and core inflation. For headline inflation, SVR achieves the lowest RMSE (0.672) compared with values exceeding 1.13 for benchmark models such as the AR, Random Walk, and VAR specifications. This represents a reduction in forecast error of approximately 40% relative to the best benchmark, highlighting the potential gains from incorporating nonlinear functional forms. GPR and tree-based methods remain broadly comparable to one another at this horizon, all improving on the linear regularized models (RMSE above 0.91 for Lasso, Ridge, and Elastic Net). For core inflation, ML models also outperform benchmarks at the 6-month horizon. SVR achieves the lowest RMSE (0.415), closely followed by GPR (0.430), representing a noticeable reduction compared to the linear regularized models and to the best benchmark specification. This difference in magnitude reflects the smoother and less volatile nature of core inflation.

Focusing on ML models across both inflation measures reveals a clear horizon-dependent structural shift. At short horizons, regularized linear models, particularly Lasso, perform well, while tree-based methods are generally weaker. As the horizon increases, SVR becomes increasingly dominant, delivering the most accurate forecasts at the 6-month horizon for both inflation measures. GPR remains competitive from medium horizons onward, particularly for core inflation. Tree-based models also improve relative to linear models by the 6-month horizon for headline inflation, suggesting that some nonlinear patterns may begin to emerge at longer horizons.

Overall, the results suggest that the relative performance of forecasting models is strongly horizon-dependent and varies across inflation measures. In the short run, linear econometric models remain difficult to outperform, reflecting the dominance of persistence in inflation dynamics. However, as the forecasting horizon increases, ML models, particularly nonlinear approaches deliver substantial gains in predictive accuracy. Similar evidence is reported by Medeiros et al. (2021) and Coulombe et al. (2020), who show that nonlinearities tend to become more relevant at longer forecasting horizons and that nonlinear machine-learning methods can improve forecasting accuracy for inflation and macroeconomic variables.

5.2 The Diebold-Mariano Test

While the RMSE provides a ranking of forecasting models, it does not indicate whether differences in forecast accuracy are statistically significant. To complement the RMSE analysis, the study therefore employ Diebold-Mariano (DM) tests to assess whether differences in out-of-sample forecast accuracy between the ML models and the corresponding benchmark models are statistically significant. Following Diebold and Mariano (1995), the results are reported in Table 7.

The Diebold-Mariano results complement RMSE results and reveal a horizon-dependent pattern. At the 1-month horizon, no ML model significantly outperforms the benchmark for either inflation measure. At the 3-month horizon, significant gains emerge for SVR and GPR in forecasting core inflation ($p < 0.05$). The evidence becomes stronger, at the 6-month horizon where all ML models significantly outperform the benchmark for core inflation ($p < 0.05$, except Ridge at ($p < 0.01$)), while SVR, GPR, Random Forest, and XGBoost achieve significant gains for headline inflation ($p < 0.01$)

5.3 Conformal Inference

To complement point forecasts, this paper employs a rolling conformal inference framework to construct prediction intervals for each forecasting model. Conformal inference provides a distribution-free approach to uncertainty quantification and does not require strong parametric assumptions regarding the distribution of forecast errors, such as normality.

A central objective of conformal prediction finite-sample marginal (unconditional) coverage, whereby the prediction set contains the realized response with a probability of at least $1 - \alpha$:

$$P\left(Y_{n+1} \in C(X_{n+1})\right) \geq 1 - \alpha,$$

where Y_{n+1} denotes the future response, X_{n+1} its corresponding feature vector, $C(X_{n+1})$ the prediction set, and $1 - \alpha$ the nominal coverage level. In the present forecasting setting, Y_{n+1} corresponds to the realized inflation rate π_{t+h}, X_{n+1} to the information available at the forecast origin t, and $C(X_{n+1})$ to the conformal prediction interval $\hat{C}_{t+h}(1 - \alpha)$.

The implementation adopts a rolling conformal procedure with a nominal coverage level of $1 - \alpha = 0.90$. At each forecast origin t, the model is estimated using a rolling training window of fixed length, while prediction intervals are calibrated using the most recent forecast errors over a fixed calibration window of size $m = 20$ months.

The finite-sample marginal coverage guarantee above relies on the exchangeability assumption, which is satisfied, in particular, when observations are independent and identically distributed (i.i.d.). This assumption generally does not hold in time-series settings due to serial dependence and time-varying volatility. The present study adopts a rolling approach, where calibration is performed using only the most recent forecast errors. This partially mitigates this limitation by allowing prediction intervals to better incorporate recent forecast performance. However, this rolling procedure does not restore the formal finite-sample coverage guarantee under exchangeability, and coverage is therefore assessed empirically over the forecasting period.

The performance of the resulting intervals is evaluated using two criteria. *Coverage* is defined as the empirical proportion of realized inflation observations that fall within the predicted interval, while *Width* is defined as the average length of the prediction interval, capturing the sharpness of the forecasts. Narrower intervals indicate higher precision, conditional on maintaining adequate coverage. Consequently, the empirical analysis focuses on the trade-off between coverage and sharpness, with particular emphasis on models that achieve tight intervals while maintaining satisfactory coverage rates.

Table 4: Conformal Inference Results for Headline Inflation

	1-Month		3-Month		6-Month	
Model	Cov. (%)	Width	Cov. (%)	Width	Cov. (%)	Width
Lasso	90.4	1.0066	84.0	2.1486	80.9	3.2546
Ridge	92.3	1.1122	94.0	2.2358	83.0	3.2720
Elastic Net	90.4	1.0356	84.0	2.1700	83.0	3.2719
SVR	90.4	1.0052	96.0	2.2458	90.0	2.2187
GPR	88.5	1.3129	90.0	2.1802	70.2	2.5754
Random Forest	90.4	2.4567	86.0	3.1394	91.5	2.6181
XGBoost	86.5	1.9344	84.0	3.3411	80.9	2.6879

Table 5: Conformal Inference Results for Core Inflation

	1-Month		3-Month		6-Month	
Model	Cov. (%)	Width	Cov. (%)	Width	Cov. (%)	Width
Lasso	90.4	0.7068	96.0	1.3636	91.5	1.6586
Ridge	92.3	0.7488	96.0	1.3136	93.6	1.8011
Elastic Net	92.3	0.7133	96.0	1.3242	91.5	1.6883
SVR	92.3	0.7937	94.0	1.2884	95.7	1.6542
GPR	90.4	0.9691	94.0	1.4017	93.6	1.5224
Random Forest	86.5	1.0275	86.0	1.8958	85.1	1.8689
XGBoost	82.7	1.3894	82.0	1.7586	85.1	2.1197

Note: Cov. denotes empirical coverage, reported as a percentage, while Width denotes the average length of the conformal prediction interval.

Tables 4 and 5 report the empirical coverage rates and average conformal prediction interval widths across forecasting horizons and ML models for headline and core inflation, respectively. Several patterns emerge from these results.

A general widening of prediction intervals is observed from the 1-month to the 3-month horizon, reflecting greater uncertainty at longer horizons. However, widths do not always increase at the 6-month horizon, indicating that the relationship between the forecasting horizon and calibrated uncertainty is not strictly increasing in these results.

For headline inflation, Ridge achieves the highest coverage at 1 month (92.3%), while SVR provides the narrowest interval (1.0052) with slightly lower coverage (90.4%). Lasso and Elastic Net show similar coverage with comparable widths, whereas GPR has lower coverage and wider intervals. Random Forest and XGBoost produce substantially wider intervals without improving coverage. At 3 months, SVR performs best, with the highest coverage (96.0%) and competitive interval width, while Ridge follows closely and tree-based models again combine wide intervals with relatively low coverage. At 6 months, SVR remains a strong performer, achieving 90.0% coverage, the second-highest after Random Forest (91.5%), with the narrowest interval overall (2.2187).

For core inflation, Ridge, Elastic Net and SVR achieve the highest coverage at 1 month (92.3%), while Lasso and Elastic Net provide the narrowest intervals among the linear models, GPR shows a markedly wider interval (0.9691) at this horizon despite adequate coverage. At 3 months, regularized linear models achieve the highest coverage (96.0%), while SVR maintains the narrowest interval overall (1.2884). At 6 months, SVR and GPR together offer the best coverage-sharpness trade-off, combining coverage close to or above nominal (95.7% and 93.6%, respectively) with intervals that are among the narrowest.

Tree-based models generally produce wider intervals than the other models, particularly at shorter horizons, though this pattern reverses for headline inflation at 6 months, where linear models yield the widest intervals. Wider intervals do not necessarily improve coverage: XGBoost frequently remains below the nominal level across both inflation measures, while Random Forest achieves the highest headline coverage at the 6-month horizon (91.5%) but shows unstable interval widths across horizons. Overall, interval width should be evaluated jointly with coverage.

Overall, the results highlight variation in the coverage-sharpness trade-off across models and forecast horizons. SVR generally offers the most favorable balance, combining coverage close to or above the nominal level with relatively narrow prediction intervals, which is also consistent with its point forecast accuracy at longer horizons reported above. GPR is competitive in several settings, particularly for core inflation, and this appears in line with its RMSE performance, where it ranks second only to SVR at the 6-month horizon for both inflation measures, nonetheless, its coverage remains more sensitive to the inflation measure and forecast horizon than SVR's. Linear models tend to generate relatively narrow intervals but can exhibit undercoverage at longer horizons, even where they achieve the lowest RMSE, as is the case for Lasso at 1- to 3-month ahead headline inflation horizons. Tree-based models tend to produce wider intervals without consistently improving coverage, alongside with their generally less accurate point forecast performance in this study. These results should be considered in the context of the specific models, sample, and forecasting exercise examined in this study.

6 SHAP-Based Forecast Decomposition

6.1 SHAP Values: Theoretical Background

To better understand the economic mechanisms underlying the forecasts generated by the ML models, this study relies on SHapley Additive exPlanations (SHAP). SHAP is an interpretability framework based on the concept of Shapley values originally developed in cooperative game theory by Shapley (1953).

In cooperative games, Shapley values provide a principled method for distributing the total payoff generated by a coalition among individual players according to their marginal contributions. Several decades later, this concept was adapted to machine learning by interpreting predictors as players and the model prediction as the payoff to be allocated among them (Štrumbelj and Kononenko, 2010). Under this framework, each explanatory variable receives a contribution reflecting its impact on a particular prediction.

SHAP decomposes this prediction as

$$\hat{\pi}_{t+h|t} = \phi_0 + \sum_{i=1}^{n} \phi_i,$$

where n denotes the total number of predictors, ϕ_0 is the baseline prediction, and ϕ_i represents the contribution of predictor i to the forecast.

The Shapley value associated with predictor i is obtained by averaging its marginal contribution across all possible subsets of predictors:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}\left(x_{S \cup \{i\}}\right) - f_S\left(x_S\right) \right],$$

where F denotes the set of all predictors, S denotes a subset (coalition) of predictors excluding i, $f_{S \cup \{i\}}$ and f_S denote the models with and without predictor i, respectively, and $x_{S \cup \{i\}}$ and x_S represent the corresponding values of the input predictors included in these models. Accordingly, each predictor's contribution is obtained by averaging its marginal contribution across the possible predictor coalitions.

Although SHAP provides considerable flexibility, its application to macroeconomic time-series data raises several methodological challenges.

The most important limitation of this approach concerns the treatment of temporal dependence. SHAP explanations may not fully account for the temporal structure of variables when evaluating feature contributions. In particular, SHAP values are computed by comparing predictions across different feature configurations relative to a background distribution, without explicitly accounting for the temporal dynamics underlying these configurations. However, current observations of macroeconomic variables are often inherently linked to their past values and to the evolution of other economic variables. Consequently, the feature combinations considered in the SHAP decomposition may not always reflect the temporal dependencies critical to the forecasting relationship. While such discrepancies may be less pronounced in some cross-sectional applications, they are particularly relevant in macroeconomic environments.

A second important challenge arises from multicollinearity. Macroeconomic variables often contain overlapping information and exhibit strong correlations. In such situations, SHAP values may distribute explanatory contributions among correlated predictors, limiting the ability to attribute importance uniquely to individual variables. Consequently, the individual contribution attributed to a particular variable should be interpreted with caution, especially when several predictors capture similar economic mechanisms.

In the context of inflation forecasting, macroeconomic variables such as interest rates, exchange rates, credit conditions, and commodity prices typically evolve jointly over time and thus contain overlapping information . As a result, SHAP explanations may rely on feature combinations that do not fully reflect the underlying economic relationships between variables, potentially producing economically implausible scenarios.

Additional challenges relate to evolving macroeconomic relationships and the definition of the baseline prediction used in the SHAP decomposition. Since the predictive relevance of inflation drivers may change over time, feature importance measures computed over extended samples may not fully capture variations in predictor contributions across different economic environments. Likewise, the interpretation of a single baseline prediction becomes less straightforward when the reference distribution used for the SHAP decomposition fails to remain representative of changing economic conditions.

To mitigate these issues, SHAP values are computed separately at each forecasting origin within the rolling forecasting framework adopted in this study. For each forecast date, the SHAP background dataset is constructed using the most recent 24 monthly observations from the corresponding training window. This approach restricts the background sample to more recent observations, thereby reducing the likelihood of generating economically implausible counterfactual combinations. In addition, the rolling framework naturally produces local explanations that evolve over time and better reflect changes in inflation dynamics across different economic regimes. Although this procedure cannot completely eliminate the challenges posed by feature dependence, multicollinearity, or other limitations inherent to SHAP-based decompositions in macroeconomic settings, it facilitates the interpretation of SHAP values in a time-series context. Therefore, SHAP values are employed in this study as an exploratory tool rather than as evidence of causal relationships between predictors and inflation.

6.2 Variable Contributions to Machine Learning Inflation Forecasts Using SHAP Values

This section presents the SHAP analysis of the model's inflation forecasts, highlighting a horizon-dependent structure in the relative contribution of explanatory variables across short-, medium-, and longer-term predictions. The corresponding SHAP decompositions are reported in Appendix E. In particular, it shows how the forecasting models differently rely on lagged inflation, monetary, financial, and external variables depending on the prediction horizon considered.

At the shortest forecasting horizon, inflation forecasts are primarily dominated by persistence effects. For both headline and core inflation, lagged inflation variables exhibit the largest SHAP contributions, indicating that recent inflation developments contain most of the predictive information for short-horizon forecasts. This finding is consistent with substantial inflation inertia. The contributions of the other selected variables remain comparatively limited, suggesting that short-term inflation forecasts rely largely on the recent history of inflation.

At the three-month horizon, the contribution structure becomes more diversified, particularly for core inflation. For core inflation, variables other than lagged inflation, including the Policy Rate, Euro Area Inflation Rate, Net Claims on Central Government, the EUR/TND exchange rate, Real Estate Price Index, and selected commodity prices, make more prominent contributions to the forecasts. For headline inflation, however, lagged inflation remains the main contributor, although some macroeconomic and external variables also play an important role. Overall, the composition of the selected variables at this horizon suggests that broader macroeconomic and external information complements inflation persistence, particularly for core inflation.

The differences in SHAP contributions across forecasting horizons become even more evident at the six-month horizon. At this stage, inflation forecasts are less dominated by persistence and reflect contributions from monetary policy, fiscal conditions, labor market developments, tourism activity, and external price factors. The Policy Rate is among the variables with the highest predictive contributions in the SHAP decomposition. This result should be interpreted as reflecting predictive importance rather than a causal monetary policy effect. Net Claims on Central Government also make substantial contributions to core inflation forecasts, which may reflect interactions between fiscal conditions and inflation dynamics.

Another notable feature at the six-month horizon is the contribution of wage dynamics. The Salaries Evolution Rate in the non-agricultural private sector contributes noticeably to the forecasts, consistent with the possible presence of second-round effects. External factors also play an important role at the six-month horizon. The Euro Area Inflation Rate, Energy Index, and global commodity prices rank among the influential predictors, highlighting the predictive relevance of external price developments imported-inflation channels in the Tunisian economy. Their importance at this horizon suggests that external conditions provide relevant information for medium-term inflation forecasts.

Taken together, the SHAP decomposition shows that the predictive information set varies across forecasting horizons. At the short horizon, both headline and core inflation forecasts are largely driven by lagged inflation. At the three-month horizon, variables beyond lagged inflation become more important, particularly for core inflation, while lagged inflation remains the dominant contributor for headline inflation forecasts despite the growing contributions of other predictors. At the six-month horizon, the contribution structure becomes more diversified, with monetary, fiscal, labor-market, real activity and external variables playing a more prominent role. These differences suggest that the information most useful for predicting inflation depends on the forecasting horizon and on the variables selected for each forecasting exercise. Nevertheless, given the methodological limitations discussed above and the fact that the explanatory variables are selected through the Sparse Group Lasso procedure, these findings should be interpreted as exploratory rather than as definitive evidence regarding the determinants of inflation forecasts. In particular, SHAP contributions reflect predictive importance and should not be interpreted as establishing causal relationships. SHAP analysis can also be used to investigate individual forecasts, providing local insights for specific prediction dates and offering additional perspective into the model's behavior over time.

7 Conclusion

While inflation remains a central focus of macroeconomic analysis, forecasting it accurately continues to be a challenging task. In this context, this paper evaluates the ability of machine learning methods to forecast Tunisian headline and core inflation at horizons of 1, 3, and 6 months ahead. Using a monthly dataset comprising macroeconomic, financial, external, and commodity price indicators, we generate forecasts and compare them with standard econometric benchmarks, including the Random Walk, AR(1), and VAR models. Beyond assessing forecasting performance, the study combines three complementary dimensions of inflation forecasting: forecast accuracy evaluation, uncertainty assessment, and an exploratory application of SHAP values to examine the contribution of explanatory variables to model predictions. The results highlight that ML models can serve as complementary forecasting approaches, providing additional predictive information alongside traditional econometric methods rather than replacing them.

The contribution of this study is therefore twofold. First, it provides a comprehensive comparison between ML and traditional econometric models for inflation forecasting in

Tunisia, while extending the analysis beyond point forecast accuracy to assess forecast uncertainty. Second, it complements the evaluation with an exploratory interpretation of ML inflation forecasts using SHAP values.

The conformal inference results complement point forecasts by providing a view of predictive uncertainty across models and forecasting horizons. Prediction intervals generally widen at longer horizons, although not monotonically. Overall, SVR offers the most favorable coverage-sharpness trade-off, particularly at longer horizons, while tree-based models tend to produce wider intervals without consistently improving coverage.

The SHAP analysis provides additional insights into how ML models rely on different predictors across forecasting horizons. At short horizons, forecasts are primarily driven by persistence effects, while monetary policy, exchange rates, fiscal variables, real economic activity variables and commodity prices become increasingly important at longer horizons. These SHAP values must nonetheless be interpreted cautiously in time-series settings, given the presence of temporal dependence and multicollinearity among predictors.

Several avenues for future research remain open, and several factors are likely to further enhance predictive performance. First, a promising direction is to extend the evaluation to longer forecasting horizons, in order to test how ML models behave as the prediction horizon lengthens further. Second, extending the sample as longer time series become available would allow more data-intensive ML models to more fully exploit nonlinear relationships. Given that the predictors selected through the Sparse Group Lasso procedure may evolve over time, future work could also consider periodically re-selecting variables and re-optimizing model hyperparameters to better capture changes in the relationships driving inflation dynamics. In particular, a larger sample would leave more room to feed models with a richer set of predictors, without relying solely on dimensionality-reduction procedures, which, while necessary given the current sample size, could be extended to incorporate a broader range of relevant predictors. A larger number of observations would also provide a more favorable setting for tree-based models when working with a richer set of predictors, by reducing the risk of overly complex tree structures relative to the available sample size. Third, incorporating alternative and higher-frequency data sources, such as news-based indicators, Central Bank communications, or social media sentiment, may contain additional forward-looking information, particularly for headline inflation, and could further improve predictive accuracy. A further avenue is to develop or adapt explainability methods for macroeconomic time-series forecasting, with the aim of mitigating the limitations of current SHAP-based approaches and providing more reliable interpretations of model predictions.

Overall, the findings suggest that ML models provide a valuable complement to traditional econometric approaches for inflation forecasting in Tunisia, Their gains are particularly pronounced at longer horizons, where nonlinearities and complex interactions become increasingly relevant. From a policy perspective, the results suggest that ML models can usefully complement, rather than replace, conventional econometric models in inflation forecasting where nonlinear relationships appear to contain particularly valuable predictive information at longer horizons.

References

Aas Kjersti, Jullum Martin, and Løland Anders (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. arXiv:1903.10464

Ahmed Nesreen, Atiya Amir, El Gayar Neamat, and El-Shishiny Hisham (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6):594–621. doi: 10.1080/07474938.2010.481556.

Angelopoulos Anastasios N., Bates Stephen (2022). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. doi: 10.48550/arXiv.2107.07511.

Angelopoulos Anastasios N., Bates Stephen, Barber Rina Foygel. (2024). Theoretical Foundations of Conformal Prediction. doi: 10.48550/arXiv.2411.11824.

Angelopoulos Anastasios N., Bates Stephen, Fannjiang Clara, Jordan Michael I., and Zrnic Tijana (2023). Prediction-powered inference. *Science*, 382:669–674. doi: 10.1126/science.adi6000

Araujo Gustavo Silva, and Gaglianone Wagner Piazza (2023). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Working Papers Series 561, Central Bank of Brazil, Research Department*, 4(2)

Artis Michael J., Banerjee Anindya, and Marcellino Massimiliano (2005). Factor forecasts for the UK. *Journal of Forecasting*, 24:279–298.

Babii Andrii, Ghysels Eric, and Striaukas Jonas (2022). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106. doi: 10.1080/07350015.2021.1899933.

Barbara Rossi, Matthieu Soupre (2017). Implementing tests for forecast evaluation in the presence of instabilities. *Stata Journal*, 17(4):850–865.

Barber Rina Foygel, Candès Emmanuel J., Ramdas Aaditya, and Tibshirani Ryan J. (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845. doi: 10.1214/23-AOS2276

Baybuza Ivan (2018). Inflation forecasting using machine learning methods. *Russian Journal of Money and Finance*, 77:42–59. doi: 10.31477/rjmf.201804.42

Breiman Leo, Friedman Jerome H., Olshen Richard A., and Stone Charles J. (1984). Classification and regression trees. Wadsworth International Group.

Chakraborty Chiranjit and Joseph Andreas (2017). Machine learning at central banks. Bank of England Working Paper No. 674.

Cortes Corinna and Vapnik Vladimir (1995). Support-vector networks. *Machine Learning*, 20(3):273–297. doi: 10.1007/BF00994018

Das Pijush Kanti and Das Prabir Kumar (2024). Forecasting and analyzing predictors of inflation rate: Using machine learning approach. *Journal of Quantitative Economics*, 22(2):493–517. doi: 10.1007/s40953-024-00384-z

Diebold Francis and Mariano Roberto (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, vol. 13 253–63.

Faust Jon and Wright Jonathan H. (2013). Forecasting inflation. In *Handbook of Economic Forecasting*, Part A, 2:2–56. doi: 10.1016/B978-0-444-53683-9.00001-3

Garcia Márcio G.P., Medeiros Marcelo C., and Vasconcelos Gabriel F.R. (2017). Real-time inflation forecasting with high dimensional models: The case of Brazil. *International Journal of Forecasting*, 33(3):679–693. doi: 10.1016/j.ijforecast.2017.02.002

Goulet Coulombe Philippe, Leroux Maxime, Stevanovic Dalibor, and Surprenant Stéphane (2022). How is Machine Learning Useful for Macroeconomic Forecasting? *Journal of Applied Econometrics*, 37(5) doi: 10.1002/jae.2910.

Hoerl Arthur E. and Kennard Robert W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67. doi: 10.2307/1267351

Inoue Atsushi and Kilian Lutz (2005). How useful is bagging in forecasting economic time series? A Case Study of U.S. CPI Inflation *Journal of the American Statistical Association*, 103(482):511–522.

Jun Maki, Hiromasa Kaneko, (2025). Benchmarking automated feature engineering in oxidative coupling of methane and the impact of domain knowledge. Results in Chemistry 18: 102730 doi:10.1016/j.rechem.2025.102730.

Kerrin Janssen and Jonny Proppe (2024). Predicting and Explaining Yields with Machine Learning for Carboxylated Azoles and Beyond. Journal of Chemical Information and Modeling, 65 (4): 1862–1872 doi: 10.1021/acs.jcim.4c02336

Lei Jing, G'Sell Max, Rinaldo Alessandro, Tibshirani Ryan J., and Wasserman Larry (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111. doi: 10.1080/01621459.2017.1307116

Liu Yang, Pan Ran, and Xu Rui (2024). Mending the crystal ball: Enhanced inflation forecasts with machine learning. IMF Working Paper.

Lundberg Scott, and Lee Su-In (2017). A unified approach to interpreting model predictions. doi: 10.48550/arXiv.1705.07874.

Lundberg, Scott & Erion, Gabriel & Chen, Hugh & DeGrave, Alex & Prutkin, Jordan & Nair, Bala & Katz, Ronit & Himmelfarb, Jonathan & Bansal, Nisha & Lee, Su-In (2019). Explainable AI for Trees: From Local Explanations to Global Understanding. 10.48550/arXiv.1905.04610.

Maehashi Kohei and Shintani Mototsugu (2020). Macroeconomic forecasting using factor models and machine learning: An Application to Japan *Journal of the Japanese and International Economies*, 58:101104. doi: 10.1016/j.jjie.2020.101104

Medeiros Marcelo C., Vasconcelos Gabriel F. R., Veiga Álvaro, and Zilberman Eduardo (2021). Forecasting inflation in a data-rich environment: The Benefits of Machine Learning Methods *Journal of Business & Economic Statistics*, 39(1):98–119.

Naghi Andrea A., O'Neill Eoghan, and Zaharieva Martina Danielova (2024). The benefits of forecasting inflation with machine learning: New evidence. *Journal of Applied Econometrics*, 39(7):1321–1331. doi: 10.1002/jae.3088

Nakamura Emi (2005). Inflation forecasting using a neural network. *Economics Letters*, 86(3):373–378.

Petar Soric, Enric Monte, Salvador Torra, Oscar Claveria (2022). Density forecasts of inflation using Gaussian process regression models University of Barcelona Working Papers 2022/10

Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, Stephane Surprenant (2020). How is Machine Learning Useful for Macroeconomic Forecasting? *University of Quebec in Montreal's School of Management* arXiv:2008.12477v1.

Rasmussen Carl E. and Williams Christopher K. I. (2005). Gaussian Processes for Machine Learning. MIT Press. doi: 10.7551/mitpress/3206.001.0001

Ribeiro Marco Tulio, Singh Sameer, and Guestrin Carlos (2016). "Why should I trust you?": Explaining the predictions of any classifier. doi: 10.1145/2939672.2939778

Rodríguez-Vargas Adolfo (2020). Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking*, 1(1). doi: 10.1016/j.latcb.2020.100012

Shah Syed Ateeb Akhter, Ishtiaq Muhammad, Qureshi Sumbal, and Fatima Kaneez (2022). Inflation forecasting for Pakistan in a data-rich environment. *The Pakistan Development Review*, 61(4):643–658.

Shapley L. S. (1953). A value for N-person games. *Contributions to the Theory of Games*, 2:307–317.

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245.

Stock James H. and Watson Mark W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*

Strumbelj Erik and Kononenko Igor (2010). An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research*, 11:1–18.

Tibshirani Robert (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Todd E. Clark, Florian Huber, Gary Koop, Massimiliano Marcellino (2022). Forecasting US Inflation Using Bayesian Nonparametric Models Federal Reserve Bank of Cleveland Working Paper No. 22-05

Vovk Vladimir, Gammerman Alex, and Shafer Glenn (2005). Algorithmic Learning in a Random World. doi: 10.1007/b106715

Yilmazkuday Hakan (2026). Forecasting Global Inflation: The Rise of Machine Learning doi: 10.2139/ssrn.7300498

Zou, H. and 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. doi: 10.1111/j.1467-9868.2005.00503.x

A Hyperparameter Tuning

Hyperparameters were selected separately for each inflation measure (headline and core) and forecast horizon (1, 3, and 6 months) using time-series cross-validation, as described in the methodology section. Table 6 summarizes the resulting hyperparameter search spaces explored for each model class.

Table 6: Hyperparameter search spaces used for model tuning.

Model	Search space
Lasso	Regularization path
Ridge	$\alpha \in \text{logspace}(-4, 4, 30)$
Elastic Net	$\alpha \in \text{logspace}(-4, 4, 30)$ ℓ_1 ratio $\in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$
SVR	$C \in [10^{-3}, 10^3]$ $\varepsilon \in [10^{-4}, 0.5]$ kernel $\in$ {linear, RBF} $\gamma \in [10^{-5}, 10]$, conditional on RBF kernel
Random Forest	n_estimators $\in [150, 250]$ max_depth $\in [2, 8]$ min_samples_split $\in [2, 10]$ min_samples_leaf $\in [2, 6]$
XGBoost	n_estimators $\in [50, 250]$ max_depth $\in [2, 4]$ learning rate $\in [0.01, 0.2]$ (log-uniform) gamma $\in [1, 4]$ reg_alpha $\in [0.1, 0.2]$ reg_lambda $\in [0.5, 2.0]$ subsample $\in [0.5, 0.9]$ colsample_bytree $\in [0.5, 0.9]$ min_child_weight $\in [1, 10]$
GPR	Kernel: Constant $\times$ Matérn ($\nu = 1.5$) + White noise Constant bounds $\in [10^{-2}, 10^2]$ Length-scale bounds $\in [10^{-2}, 10^2]$ Noise-level bounds $\in [0.5, 1]$ (Core); $[0.9, 5]$ (Headline)

Note: All model optimizations were performed using five-fold time-series cross-validation to preserve the temporal ordering of the observations. The linear models (Lasso, Ridge, and Elastic Net) were tuned using their respective native cross-validation routines (LassoCV, RidgeCV, and ElasticNetCV). SVR and tree-based models were optimized using BayesSearchCV with RMSE as the selection criterion over 40 iterations. For GPR, the kernel parameters were calibrated using the L-BFGS-B optimizer with 10 optimizer restarts.

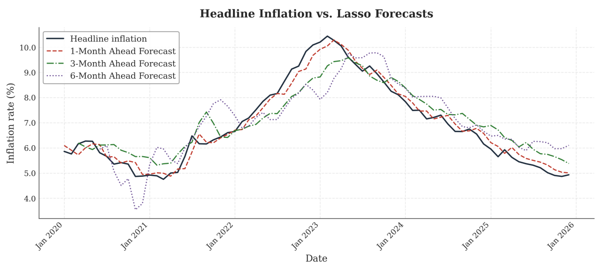
B SHAP Implementation Details

This appendix describes the implementation of the SHAP analysis presented in the main text. For each inflation measure and forecast horizon, SHAP values were computed for the best-performing model identified in Table 2 and Table 3, within the same rolling out-of-sample framework used for forecast evaluation.

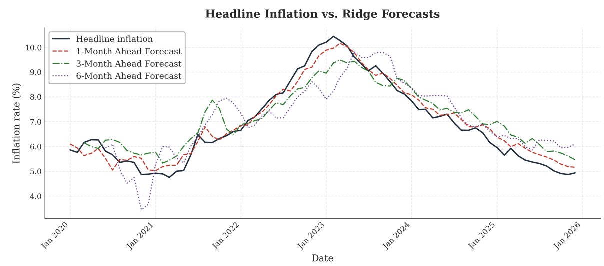
The choice of explainer depends on the linearity of the underlying model: for linear specifications, SHAP values are computed using `LinearExplainer`; for nonlinear specifications, SHAP values are approximated using `KernelExplainer`.

C Projected vs. Actual inflation

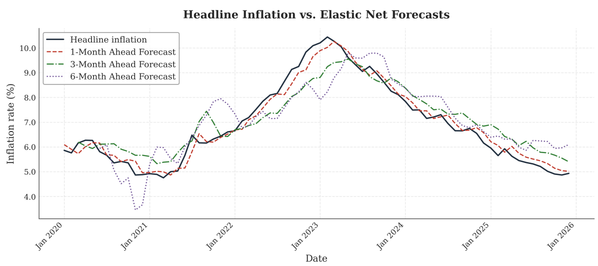
C.1 Headline Inflation



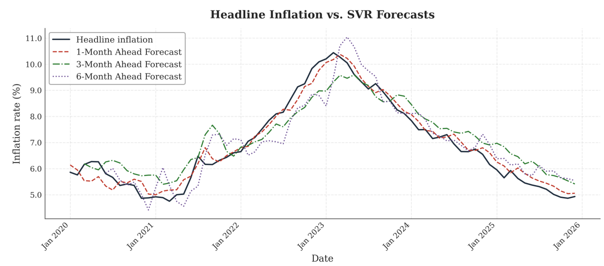
(a) Lasso



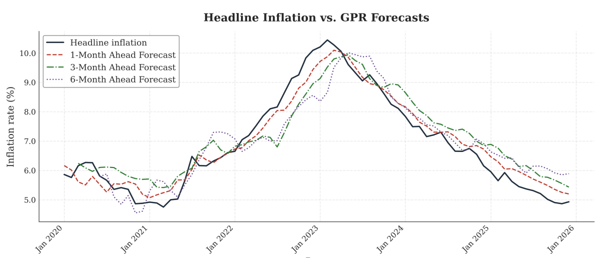
(b) Ridge



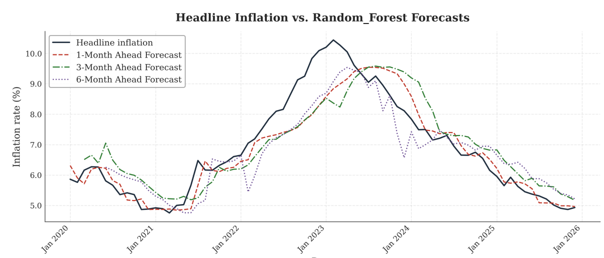
(c) Elastic Net



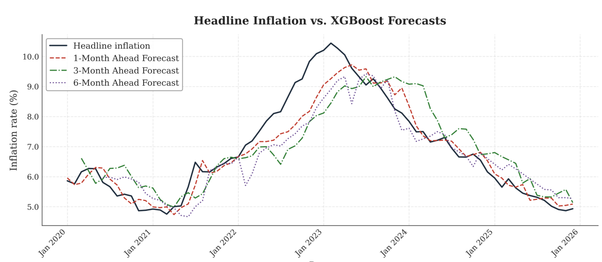
(d) SVR



(e) GPR



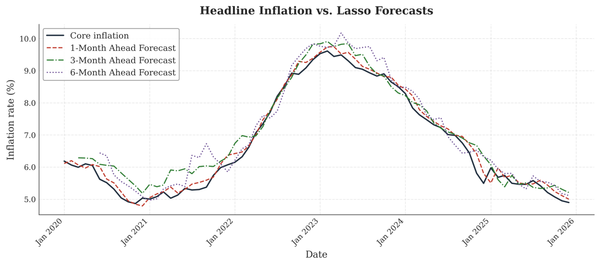
(f) Random Forest



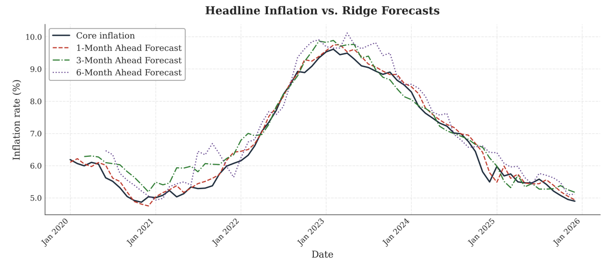
(g) XGBoost

Figure 2: Out-of-sample forecasts of headline inflation across models

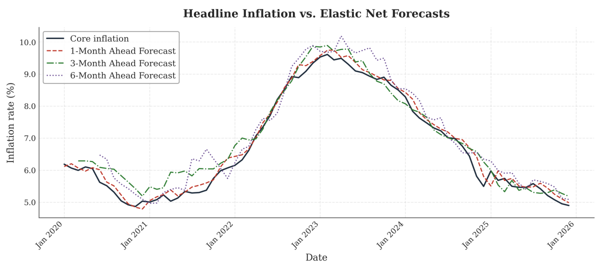
C.2 Core Inflation



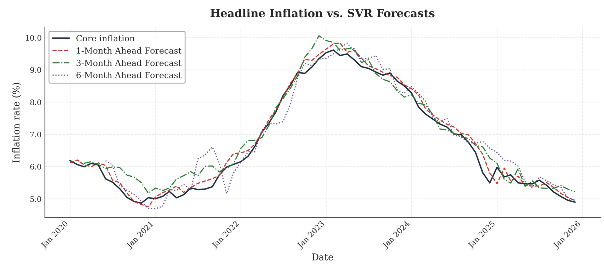
(a) Lasso



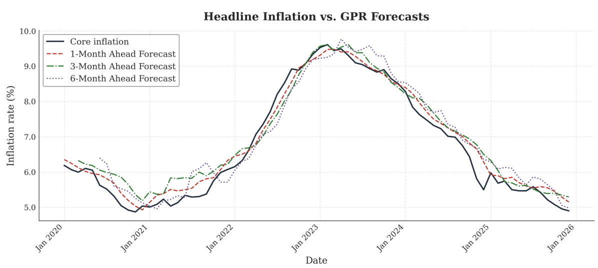
(b) Ridge



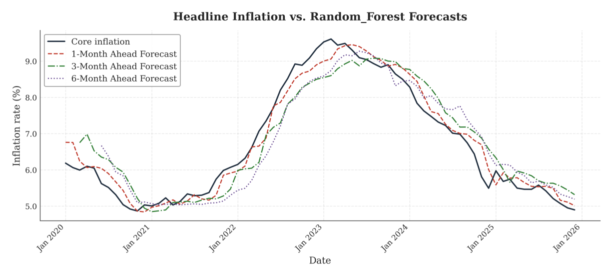
(c) Elastic Net



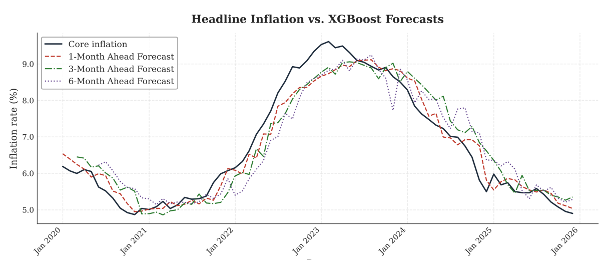
(d) SVR



(e) GPR



(f) Random Forest



(g) XGBoost

Figure 3: Out-of-sample forecasts of core inflation across models

D Diebold Mariano (1995) test

Table 7: Diebold-Mariano test results: ML models versus the corresponding benchmark

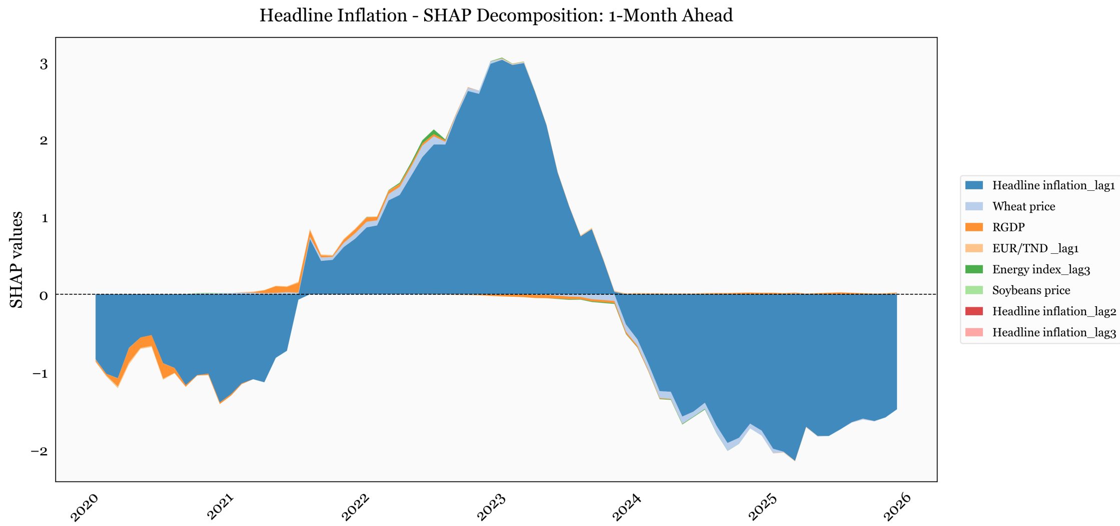
Headline	h = 1		h = 3		h = 6	
Model	DM stat.	p-value	DM stat.	p-value	DM stat.	p-value
LASSO	3.260***	0.0017	0.639	0.5248	−1.759*	0.0832
Ridge	4.071***	< 0.001	0.998	0.3217	−1.745*	0.0856
ElasticNet	3.371***	0.0012	0.712	0.4790	−1.746*	0.0855
SVR	3.165***	0.0023	1.029	0.3072	−3.673***	< 0.001
GPR	5.744***	< 0.001	0.900	0.3715	−3.139***	0.0025
RandomForest	4.003***	< 0.001	1.762*	0.0825	−4.539***	< 0.001
XGBoost	3.701***	< 0.001	1.791*	0.0777	−2.818***	0.0064

Core	h = 1		h = 3		h = 6	
Model	DM stat.	p-value	DM stat.	p-value	DM stat.	p-value
LASSO	−0.351	0.7264	−1.639	0.1057	−2.022**	0.0473
Ridge	−0.140	0.8889	−1.597	0.1148	−1.971*	0.0529
ElasticNet	−0.266	0.7910	−1.601	0.1140	−1.997**	0.0499
SVR	0.051	0.9598	−2.174**	0.0331	−2.179**	0.0329
GPR	2.504**	0.0146	−2.130**	0.0367	−2.260**	0.0271
RandomForest	4.045***	< 0.001	0.422	0.6741	−2.347**	0.0219
XGBoost	4.474***	< 0.001	−0.532	0.5964	−2.435**	0.0176

Notes: Entries report the Diebold-Mariano (1995) test statistic for the null hypothesis of equal predictive accuracy between each ML model and the corresponding benchmark, using squared-error loss. The test is conducted with the appropriate forecast horizon, (h=1,3,6). the Diebold-Mariano test statistic is reported with the Harvey-Leybourne-Newbold (1997) small-sample correction. The benchmark corresponds to the best-performing benchmark selected for each inflation measure and forecast horizon. A negative statistic indicates that the ML model has lower average squared loss than the corresponding benchmark, while a positive statistic indicates the reverse. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

E SHAP Decomposition of Headline and Core Inflation

Note: SHAP values represent each variable's contribution to the model's inflation forecast at each forecasting date.



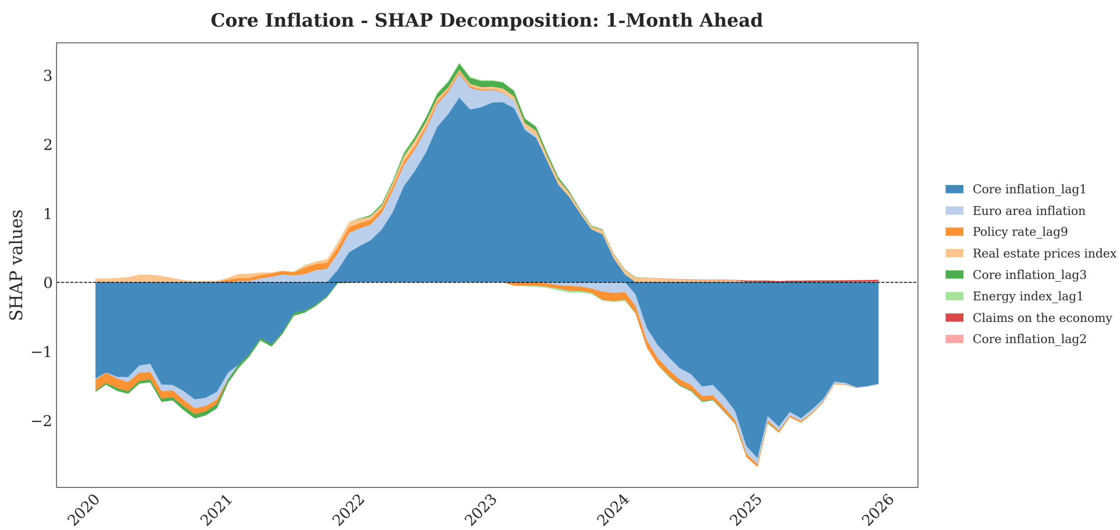


Figure 4: SHAP Decomposition – 1-Month Ahead.

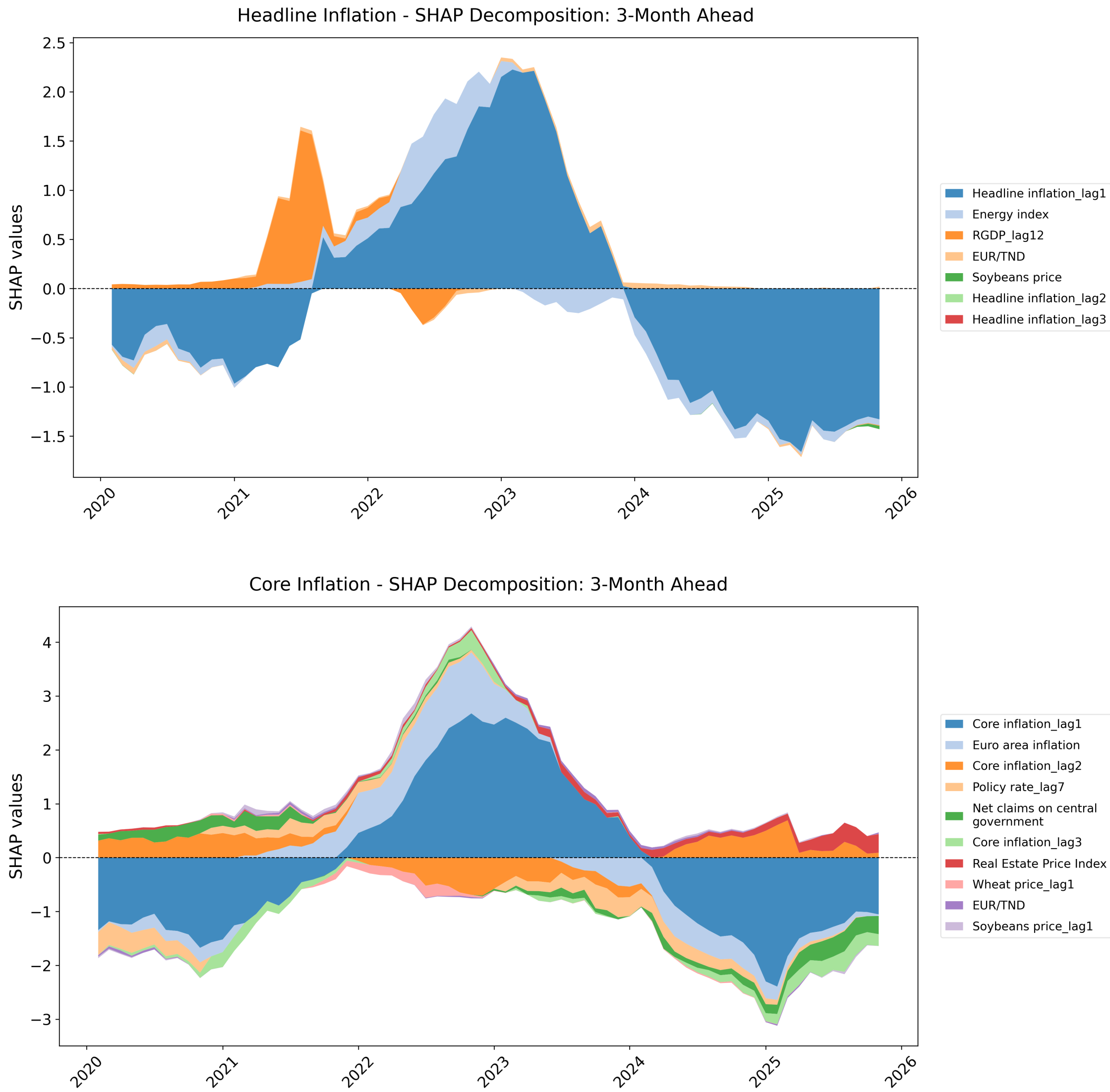
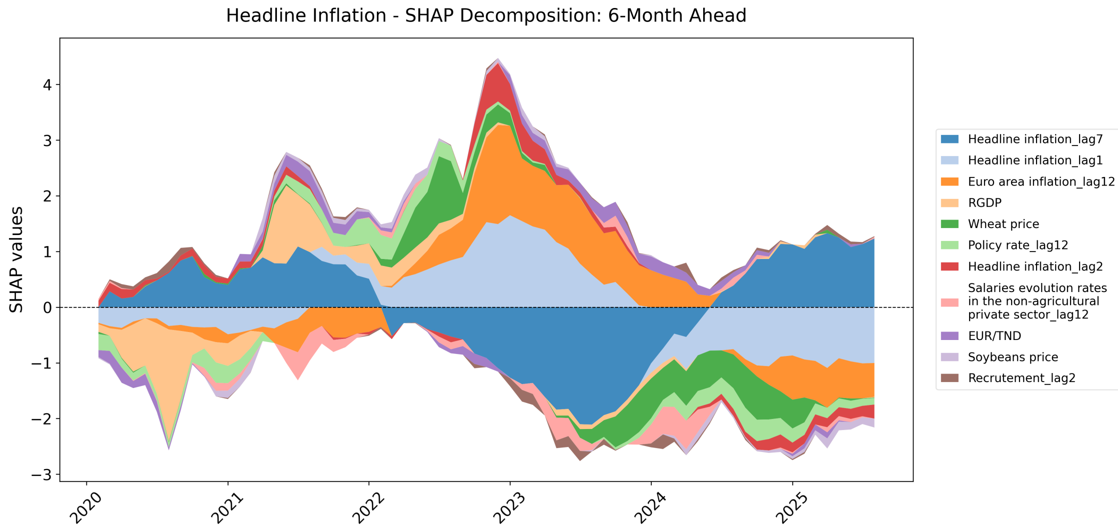


Figure 5: SHAP Decomposition – 3-Month Ahead.



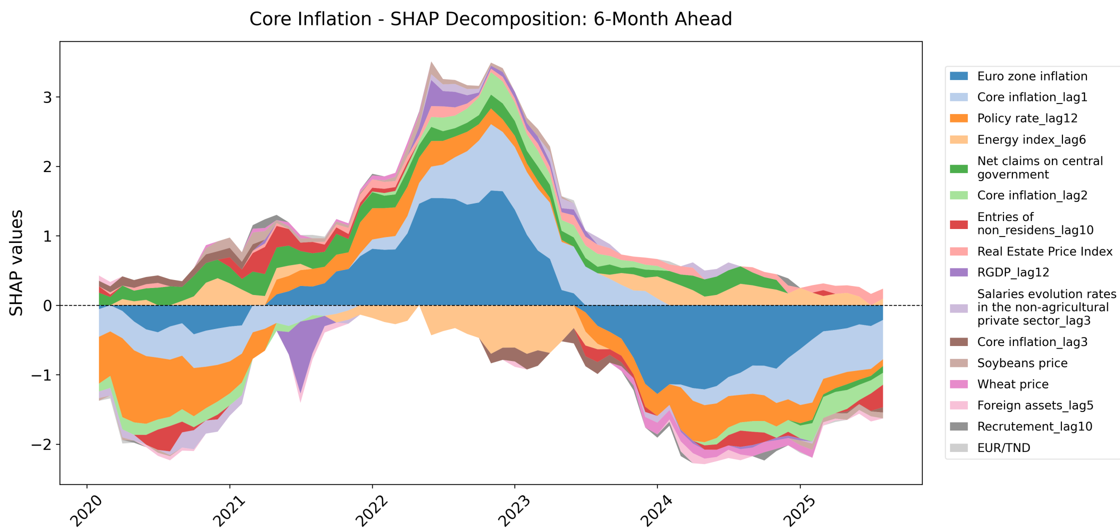


Figure 6: SHAP Decomposition – 6-Month Ahead.