BANCO CENTRAL DE COSTA RICA

DOCUMENTOS DE TRABAJO
N.º 06 | 2026

Métodos de aprendizaje automático para la predicción de la inflación en múltiples horizontes: un análisis para Costa Rica.

Manuel Esteban Sánchez Gómez

Fotografía de portada: “Presentes”, conjunto escultórico en bronce, año 1983, del artista costarricense Fernando Calvo Sánchez. Colección del Banco Central de Costa Rica.

Métodos de aprendizaje automático para la predicción de la inflación en múltiples horizontes: un análisis para Costa Rica.

Manuel Esteban Sánchez Gómez‡

Resumen

Este documento evalúa el desempeño de los métodos de aprendizaje automático (ML) para pronosticar la inflación interanual en Costa Rica, utilizando datos mensuales de 2012 a 2025 y comparándolos con estándares de referencia dentro de un esquema de muestra móvil fuera de muestra. Las técnicas de ML son especialmente útiles para captar no linealidades e interacciones complejas entre la inflación y un amplio conjunto de variables macroeconómicas. Los resultados muestran que los métodos de ensamblaje no lineal como XGBoost y BART presentan los mayores beneficios en horizontes cortos, mientras que los métodos de reducción lineal son más competitivos en horizontes largos.

Palabras clave: Inflación, Pronósticos, Machine Learning, Política Monetaria, Modelos Híbridos.
Clasificación JEL.: C53, C55, E37.

‡ División Económica, BCCR. Email: sanchezgm@bccr.fi.cr

Machine Learning Methods for Multi-Horizon Inflation Forecasting: A Comparative Analysis for Costa Rica.

Manuel Esteban Sánchez Gómez[‡]

The ideas expressed in this paper are those of the authors and not necessarily represent the view of the Central Bank of Costa Rica.

Abstract

This paper evaluates the performance of machine learning (ML) methods for forecasting year-over-year inflation in Costa Rica using monthly data from 2012-2025 and compares their performance against standard benchmarks within a rolling out-of-sample framework. ML techniques are particularly useful for capturing nonlinearities and complex interactions between inflation and a broad set of macroeconomic covariates. The results show that nonlinear ensemble methods such as XGBoost and BART provide the strongest gains at short horizons, while linear shrinkage methods are more competitive at longer horizons.

Key words: Inflation, Macroeconomic Forecasting, Machine Learning, Monetary Policy, Hybrid Models.
JEL codes: C53, C55, E37.

[‡]Economic Division, BCCR. Email address sanchezgm@bccr.fi.cr

Contents

1 Introduction

Price stability, reflected in a low and stable inflation rate, is the main objective of the Central Bank of Costa Rica and is fundamental for achieving sustained economic development and social well-being. High or volatile inflation creates uncertainty, affects household purchasing power, distorts investment decisions, and complicates financial planning for both the public and private sectors. Therefore, having accurate inflation forecasts is essential for conducting monetary policy and making informed decisions.

In recent years, advances in data science and the development of machine learning (ML) algorithms have opened new possibilities for analyzing and predicting complex economic phenomena. Unlike traditional econometric methods, which often require strict assumptions about the relationship between variables, ML approaches allow for flexible identification and modeling of nonlinearities and hidden patterns in the data. This capability is especially valuable in the Costa Rican inflationary context, characterized by the influence of multiple internal and external factors that interact dynamically.

This paper explores the use of various ML techniques—including random forests, XGBoost, Elastic Net, Ridge and Lasso Regressions, Bayesian Additive Regression Trees (BART) and a hybrid model that combines a Vector Auto-regression model (VAR Model) with a BART to forecast year-over-year inflation in Costa Rica using data from 2012 to 2023 for model training and evaluating their performance with a one step ahead rolling window strategy. These methods will be implemented and compared with classical benchmarks, such as a random walk and AR(1) models to evaluate their predictive performance and the practical utility of their results for monetary management. Additionally, we used the Diebold and Mariano Test to check whether these models present statistically significant predictive ability compared to the benchmarks. The implementation of these models aims to improve the accuracy of forecasts and provide more robust tools for the Central Bank, thereby contributing to price stability and institutional credibility.

To complete this task, we developed a database of 172 macroeconomic variables with monthly frequency. The variables included in this dataset are selected according to what macroeconomic theory suggests are the determinants of inflation dynamics, such as input prices and costs, aggregate demand, public finances, expectations, labor markets and financial conditions. The integration of these variables combined with ML techniques allows us to discover complex non-linear hidden interactions and to anticipate possible changes in inflation trends, which strengthen the Central Bank's capacity to proactively respond to deviations of inflation from its target.

Our results suggest that i) all multivariate models outperform naive persistence benchmarks, highlighting the informational content of macroeconomic covariates; ii) the nonlinear ensemble methods dominate at short horizons (up to (h=6), suggesting the presence of nonlinearities that are relevant in the inflation dynamics, and iii) the linear models regain competitiveness at longer horizons, where overfitting risks and parameter instability might increase. Also, the Diebold-Mariano tests show that among all the models evaluated, XGBoost and BART in short horizons (less than 6 months) presents the highest predictive ability.

This paper contributes to the economic literature in three ways. First, to the best of our knowledge, it provides the first systematic evaluation of machine learning methods for inflation forecasting in Costa Rica under a realistic real-time information environment. Second, it assesses the forecasting performance of a residual-hybrid VAR+BART strategy in an inflation forecasting context. Third, it documents horizon-specific differences in predictive performance, showing that nonlinear methods dominate at short horizons while linear shrinkage methods regain competitiveness at longer horizons.

In summary, the use of ML techniques for inflation forecasting represents a significant methodological advance that can transform the way monetary policy is approached in Costa Rica, strengthening analytical tools and improving decision-making in an increasingly uncertain and changing environment. It is worth mentioning that the implementation of these methods has gained relevance as a practice in central banks of both emerging and advanced economies.

This paper is organized as follows. Section 2 analyzes the evolution and main stylized facts of inflation in Costa Rica over the period 2012–2025. Section 3 reviews the increasing literature on the application of ML methods to macroeconomic forecasting. Sections 4 and 5 describe the dataset and outline the empirical strategy, including the methodological framework and the specific ML models implemented. Section 6 presents the main empirical results, and Section 7 concludes.

2 Stylized facts of Inflation in Costa Rica

From 2012 to 2025, inflation in Costa Rica has exhibited notable shifts, shaped by both external and domestic factors as well as significant changes in monetary policy frameworks. Figure 1 shows the dynamics of the inflation rate, monetary policy rate and the monthly index of economic activity (IMAE).

In the early part of the decade, inflation rates were relatively volatile, influenced by international commodity prices, exchange rate adjustments, and fluctuations in local economic activity. The Central Bank of Costa Rica closely managed monetary conditions, aiming to balance price stability and economic growth amid these challenges.

A key turning point occurred in 2015, when Costa Rica officially transitioned to a floating exchange rate regime. This change allowed the local currency to respond more flexibly to market dynamics, enabling monetary policy to operate with greater independence. This change of regime helped to moderate inflationary pressures over time, as the exchange rate became a buffer against external shocks, and the Central Bank gained additional capacity to guide inflation expectations.

In 2018, the country marked another important milestone with the formal adoption of the inflation targeting framework. This shift established a clear commitment to maintaining inflation within a specific target range, anchored by enhanced transparency and communication from the Central Bank. Since then, inflation dynamics have generally become more predictable and moderate, with rates staying relatively closer to the inflation target, (3 percent with a tolerance range of 1 percentage point), which suggests an improvement in credibility of the monetary authority.

The recent years have been challenging for achieving of the Central Bank's target, as global events and domestic factors have at times tested the resilience of the framework, requiring ongoing vigilance and adaptation by policymakers. Particularly, while the COVID-19 shock implied a reduction in inflationary pressures in 2020, the subsequent increase in transportation costs and raw materials prices led to a significant increase in inflation which was addressed through several adjustments in monetary policy rate.

Later, as the external shock reversed, inflationary pressures eased. Commodity prices, in particular, followed a lower-than-expected path, while domestic supply shocks associated with weather conditions and their impact on food prices have helped keep inflation below the target range, even pushing it into negative territory.

In this context, output growth after the pandemic has maintained annual rates above 4 percent, consistent with the latest estimates of long-run output growth (around 3.8 percent) according to Rodríguez-Vargas (2022).

In sum, the recent years have been depicted by a sequence of consecutive and unusual shocks that not only increase uncertainty but also reinforce the challenge of forecasting inflation. In that sense, traditional econometric models may struggle to capture the complex, nonlinear relationships and rapid shifts that characterize today's economic environment. ML approaches, by contrast, can adapt to changing patterns in the data, incorporate a wide range of variables and lags, and improve the accuracy and timeliness of inflation predictions. This flexibility and robustness help policymakers, businesses, and investors make more informed decisions amid evolving risks and uncertainties.

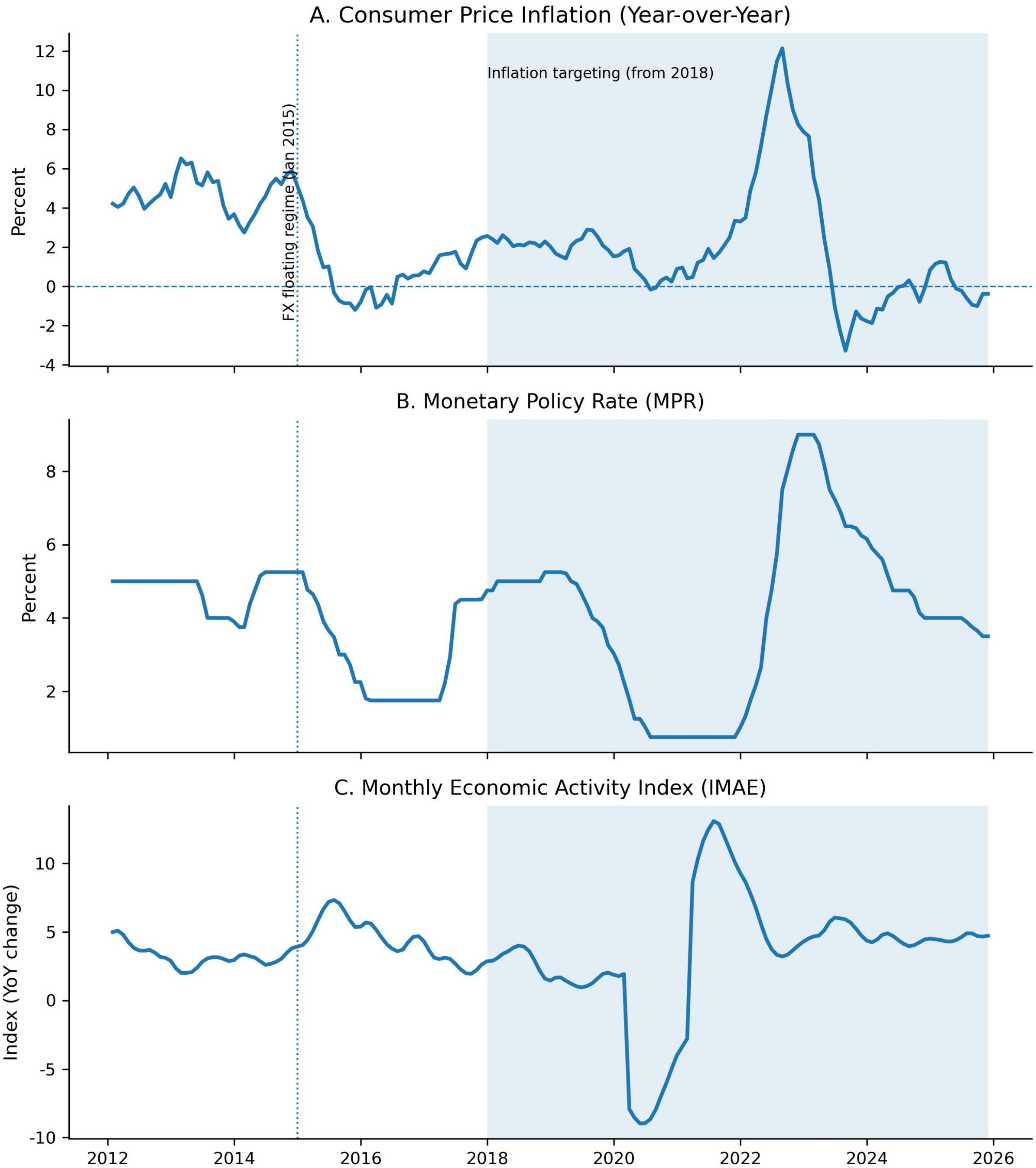


Figure 1: Consumer price inflation (year-over-year). The shaded area denotes the inflation-targeting regime (from 2018 onward). The vertical line marks the adoption of the floating exchange rate regime (January 2015).

3 Literature Review

Literature regarding traditional methods of econometrics to forecast macroeconomic variables such as inflation is vast and typically uses linear time series models based on theoretical relationships among variables and states for appropriate properties of the estimators, a comprehensive analysis of some of these methods and its performance can be found on Faust and Wright (2013).

However, our main interest for this paper is on applications of machine learning algorithms to produce forecasts, which can be non-linear, are not ruled by theoretical relationships among variables and has become very popular in the last years for its performance, especially on periods of high uncertainty. In this section we present some research related to this techniques applied to macroeconomic forecasting, with special attention to those focused on inflation.

Inoue and Kilian (2008) use bagging techniques to forecast inflation and with positive results in comparison to its benchmarks. Medeiros and Vasconcelos (2016) show that the use of high dimensional data models produce on average small forecasting errors for macroeconomic variables. Later, Medeiros et al. (2021) test several models to forecast inflation for United States and find that random forest deserves particular attention, as it presents the best performance to forecast inflation.

Abraham et al. (2022) tested 73,884 model specifications by using a database of 1.180 variables to predict 188 quarterly revenue and expenditure series for the services sector as published in the Quarterly Service Survey (QSS), a key data set that accounts for nearly 80 percent of the revisions to Personal Consumption Expenditure for Services (PCE Services). The authors state that ensemble methods such as Random Forests afford the highest chance of reducing revisions.

Nyman and Ormerod (2017) use a small set of explanatory variables from financial markets in United Kingdom and train a Random Forest algorithm to predict recessions; while the results are promising it is worth to note that the authors used the default input parameters. Similarly, Cicceri et al. (2020) and Tehranian (2023) trained models for Italy and The United States respectively.

Araujo and Gaglianone (2023) estimate more than fifty models to forecast inflation in Brazil and finds that ML techniques such as random forest or XGBoost outperform traditional econometric models.

Huber et al. (2023) argue that, conventional linear models work poorly during extreme episodes such as the pandemic due to their inability to effectively deal with the extreme observations that have occurred, while ML models exhibit advantages. The authors developed a mixed model between a VAR and a Bayesian additive regression trees for nowcasting GDP growth.

Magazzino et al. (2025) calibrates neural networks to forecast the unemployment rate for developed economies over the period 1998 to 2016. The authors finds significant improvements in accuracy of unemployment forecasting for the selected countries, allowing the analysis of the contributions of each input to unemployment estimation as well.

Goulet Coulombe et al. (2021) assess the forecasting performance of a variety of standard and ML forecasting methods for key UK economic variables, with a special focus on the COVID-19. Some of the methods evaluated are penalized regressions, boosted trees and random forests, Kernel Ridge Regression, Neural Networks, and a Macroeconomic Random Forest (MRF), which uses a linear part within the leafs, and Moving Average Rotation of X (MARX), a feature engineering technique which generates an implicit shrinkage more appropriate for time series data. One of the main results is that ML methods provide substantial gains when short-term forecasting several indicators of the UK economy.

Finally, Liu (2024) estimated forecast for core inflation of Japan's economy in the post-pandemic period. The author found that Lasso regressions gives the best performance and argues that ML models improve forecasting performance by incorporating a wider range of variables, allowing for non-linear relationships, and focusing on out-of-sample performance.

4 Data

A database was built based on time series obtained from sources such as the Central Bank of Costa Rica, the National Census and Statistics Institute, and the Ministerio de Hacienda. The database contains macroeconomic series that are theoretically related to inflation, such as price and cost variables, economic activity indicators, inflation expectations, labor market, public finances, and financial conditions. Table 1 provides a detailed list of the 43 variables included. Additionally, lags of 1, 3, 6, and 12 months were created for each variable, resulting in a dataset with a total of 172 columns containing monthly information for the period from January 2013 to December 2025, accounting for 28,896 observations in total.

Almost all variables are expressed as year-over-year growth rates, except for interest rates and some labor market indicators, which are shown as percentages, and public finance variables which are included as a proportion of Gross Domestic Product. Table A1 presents the descriptive statistics for each variable in the dataset.

Table 1: Variables and definitions

Code	**Variable**	**Unit / Transformation**	**Source**
inf_int	Consumer Price Index (CPI)	YoY change	INEC
ipp_man	Manufacturing Producer Price Index (PPI)	YoY change	BCCR
ip_mpi	Commodity (raw materials) Price Index	YoY change	BCCR
wti	International oil price (WTI)	YoY change	BCCR
ismn	Nominal minimum wage index	YoY change	BCCR
ismr	Real minimum wage index	YoY change	BCCR
w_nom_tc	Household nominal income	YoY change	INEC
w_real_tc	Household real income	YoY change	INEC
imae	Monthly Economic Activity Index (IMAE)	YoY change	BCCR
imae_re	IMAE — special regime	YoY change	BCCR
imae_rd	IMAE — definitive regime	YoY change	BCCR
imae_sd	IMAE — seasonally adjusted series	YoY change	BCCR
exp_rd	Goods exports — definitive regime	YoY change	BCCR
exp_re	Goods exports — special regime	YoY change	BCCR
exp_tot	Total goods exports	YoY change	BCCR
imp_rd	Goods imports — definitive regime	YoY change	BCCR
imp_re	Goods imports — special regime	YoY change	BCCR
imp_tot	Total goods imports	YoY change	BCCR
cred_mn	Private-sector credit in local currency	YoY change	BCCR
cred_me	Private-sector credit in foreign currency	YoY change	BCCR
cred_tot	Total private-sector credit	YoY change	BCCR
exp_inf_12_prom	12-month-ahead infl. expectations (survey, average)	Percent	BCCR
exp_inf_12_med	12-month-ahead infl. expectations (survey, market)	Percent	BCCR
exp_inf_12_merc	12-month-ahead infl. expectations (market, average)	Percent	BCCR
brecha_exp_enc	Inflation expectations gap (survey-based)	Percent	BCCR
brecha_exp_mer	Inflation expectations gap (market-based)	Percent	BCCR
exp_inf_12_enc_var	12-month-ahead inflation expectations (survey, average)	Change (pp)	BCCR
exp_inf_12_merc_var	12-month-ahead inflation expectations (survey, market)	Change (pp)	BCCR
tnp	Net labor force participation rate	Percent	INEC

Continued on next page

Code	Variable	Unit / Transformation	Source
to	Employment rate	Percent	INEC
ti	Inactivity rate	Percent	INEC
td	Unemployment rate	Percent	INEC
ts	Underemployment rate	Percent	INEC
res_fin_gc	Government overall financial balance	Share of GDP	MH
deuda_tot	Government total debt	Share of GDP	MH
deuda_int	Government domestic debt	Share of GDP	MH
tbp	Basic Passive Rate (TBP)	Percent	BCCR
tpm	Monetary Policy Rate (MPR)	Percent	BCCR
prime	Prime Rate	Percent	BCCR
tcn_var	Nominal exchange rate	YoY change	BCCR
basem_var	Monetary base	YoY change	BCCR
m1_var	Money supply (M1)	YoY change	BCCR
itcer_var	Real exchange rate index (ITCER)	YoY change	BCCR

Notes: "YoY change" denotes year-over-year percentage change. "Percent" denotes levels in percentage points. "Change (pp)" denotes first differences expressed in percentage points.
Source: Authors' compilation based on Instituto Nacional de Estadística y Censos (INEC), Banco Central de Costa Rica (BCCR), and Ministerio de Hacienda (MH).

5 Methodology

This section describes the procedure used to evaluate a set of machine-learning and Bayesian nonparametric models for inflation projections at multiple horizons, similar to the approach followed by Medeiros et al. (2021) and Liu (2024). The same protocol is applied to all the models implemented: Extreme Gradient Boosting (XGB), Random Forests (RF) and penalized linear regressions (Ridge, Elastic Net, and Lasso) to ensure comparability, except for the Bayesian Additive Regression Trees (BART), and the hybrid model Bayesian Additive Regression Trees with Vector Autoregressive Regressions (VAR+BART) as we used the hyper-parameters suggested by Chipman et al. (2010).[1]

5.1 Model, target variable and forecast horizons

Suppose a model for inflation such as:

$$\pi_{t+h} = F_h(x_t) + u_{t+h}, \qquad h = 1, \dots, H, \qquad t = 1, \dots, T, \tag{1}$$

where π_{t+h} is the inflation (year-over-year) in month $t+h$; $x_t = (x_{1t}, \dots, x_{nt})'$ is an n-vector of covariates that also contains lags of π_t and the rest of the covariates; $F_h(\cdot)$ is a function that can be a single model or an ensemble of different specifications, linear or non-linear. This function maps the relation between the covariates and inflation in $t+h$; and u_t is a zero-mean random error. The forecast horizon $h \in \{1, 3, 6, 12\}$ (in months).

Then, the objective is to produce an h-step-ahead forecast

$$\widehat{\pi}_{t+h|t} = \widehat{F}_{h,\, t-R_h+1:t}(x_t). \tag{2}$$

[1] BART is estimated with default regularization settings, which are computationally efficient and typically robust; hence, results may be conservative relative to extensive tuning.

Where $\widehat{F}_{h,\,t-R_h+1:t}(x_t)$ is the estimated value for the function evaluated on data from time $t - R_h + 1$ up to t, and R_h is the window size. We use direct forecasts, since we don't generate any predictions for lagged covariates.

5.2 Information set, data availability, and feature construction

Macroeconomic predictors are observed with publication lags. To mimic a realistic information set, predictors are grouped by their typical availability:

- **Group t:** variables available contemporaneously at time t (e.g., financial indicators, expectations, external prices).
- **Group $t-1$:** variables typically available with a one-month delay (e.g., activity and credit aggregates).
- **Group $t-2$:** variables typically available with a two-month delay (e.g., fiscal indicators).

Let $x_{j,t}$ denote predictor j at time t, and let $s_j \in \{0, 1, 2\}$ be its availability shift (0 for Group t, 1 for Group $t-1$, and 2 for Group $t-2$). For each predictor, a set of lags $\mathcal{L} = \{1, 3, 6, 12\}$ is created. The effective lag used in estimation is

$$\ell^* = \ell + s_j, \qquad \ell \in \mathcal{L}. \tag{3}$$

Hence, the model uses $x_{j,t-\ell^*}$ to ensure that only information available at time t enters $\mathcal{I}_t$.

In addition, the contemporaneous level $x_{j,t}$ is included only for predictors in Group t. The feature vector for horizon h is denoted $X_t^{(h)}$, which stacks all contemporaneous predictors (when available) and the lagged predictors selected for that horizon. To allow for short-run inflation dynamics, the first difference of inflation is also considered:

$$\Delta\pi_t = \pi_t - \pi_{t-1}, \tag{4}$$

and its lags $\Delta\pi_{t-\ell}$, $\ell \in \mathcal{L}$, is included for medium and long horizons (e.g., $h \in \{6, 12\}$).

5.3 Train–test split and evaluation period

The sample is divided into two tables, one for training or calibration which includes observations from 2012 to 2023, and a testing table with data for 2024 to 2025; where forecasts are evaluated using a rolling moving-window of 90 months scheme.

The implementation of a rolling window estimation framework in inflation forecasting contributes to mitigate the risks associated with parameter instability and structural breaks. This strategy prioritizes recent economic dynamics while discarding obsolete historical information that may no longer reflect the current monetary regime.

Then, for each horizon h, the learning target aligns with the h-step-ahead outcome,

$$y_t^{(h)} = \pi_{t+h}. \tag{5}$$

5.4 Time-series cross-validation for hyperparameter tuning

The hyper-parameters are selected using time-series cross-validation (CV) within the training period. The tuning sample for horizon h contains N_h usable observations after aligning $y_t^{(h)}$ and removing missing values. We construct 3 sequential splits with fixed test-block size 12 and expanding training sets. For split $k \in \{1, \ldots, K\}$, define:

$$\mathcal{T}_k = \{1, \ldots, n_k\}, \tag{6}$$

$$\mathcal{V}_k = \{n_k + 1, \ldots, n_k + m\}, \tag{7}$$

where n_k increases with k. This preserves temporal ordering and prevents the use of future data to forecast past data. The loss for a candidate hyperparameter vector θ is computed as an average across splits:

$$\text{CVLoss}^{(h)}(\theta) = \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}\left(y_{\mathcal{V}_k}^{(h)}, \widehat{y}_{\mathcal{V}_k}^{(h)}(\theta)\right), \tag{8}$$

where $\mathcal{L}(\cdot)$ is the root mean squared error (RMSE) computed on the validation block,

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{t \in \mathcal{V}_k} \left(\pi_{t+h} - \widehat{\pi}_{t+h|t}\right)^2}. \tag{9}$$

The selected hyperparameters satisfy $\hat{\theta}^{(h)} = \arg\min_\theta \text{CVLoss}^{(h)}(\theta)$.

5.5 Bayesian optimization of hyperparameters

Hyperparameters are tuned via Bayesian optimization (BO), a sophisticated framework designed to optimize objective functions that are computationally expensive. It operates by constructing a probabilistic surrogate model—typically a Gaussian Process—to approximate the unknown function based on historical evaluations. The search process is guided by the Expected Improvement (EI) acquisition function, which mathematically evaluates the anticipated gain over the current best observed value. By strategically balancing the exploration of high-uncertainty regions with the exploitation of promising areas, this approach identifies the global optimum using significantly fewer iterations than traditional grid or random search methods. A detailed explanation of this method can be found in Snoek et al. (2012). To implement this method we used `BayesSearchCV` from `scikit-optimize`.

Particularly, BO treats the CV objective $\text{CVLoss}^{(h)}(\theta)$ as a function and iteratively proposes hyperparameter candidates by fitting a probabilistic surrogate model and maximizing an acquisition function:

$$\theta_{t+1} = \arg\max_{\theta \in \Theta} a_t(\theta), \tag{10}$$

where $a_t(\theta)$ trades off exploration (uncertainty) and exploitation (low predicted loss). This approach is more sample-efficient than grid or random search when the hyperparameter space is high-dimensional.

In this paper, we tuned the hyperparameters to estimate models to forecast inflation for 1, 3, 6 and 12 months ahead. Additionally, we implement several machine learning methods, such as Random Forest, Extreme Gradient Boosting, Elastic Net, Ridge and Lasso Regressions. Table A2 details the hyperparameters estimated in each case.

5.6 Out-of-sample evaluation with a rolling window

Forecast performance is assessed using a rolling window of fixed length $W = 90$ months. For each test origin t (starting in January 2024) and horizon h, define the estimation window:

$$\mathcal{W}_t = \{t - W, \ldots, t - 1\}. \tag{11}$$

At each origin t, the model is re-estimated on $\mathcal{W}_t$ using the horizon-specific features $X_\tau^{(h)}$ and targets $\pi_{\tau+h}$ for $\tau \in \mathcal{W}_t$, and a forecast $\widehat{\pi}_{t+h|t}$ is produced. This strategy mimics operational forecasting and allows the relationship between inflation and predictors to vary over time.

5.7 Benchmarks

A common practice in "horse race papers" such as this one, is to compare the forecasts obtained from the different models with simple benchmarks to contextualize their predictive accuracy:

- **Random Walk (RW):** $\widehat{\pi}^{\mathrm{RW}}_{t+h|t} = \pi_t$
- **Autoregressive benchmark AR(1):** a linear model

$$\pi_t = c + \phi\,\pi_{t-1} + u_t, \tag{12}$$

estimated within each rolling window, producing $\widehat{\pi}^{\mathrm{AR}(1)}_{t+h|t}$ by iterating the one-step-ahead projection (or, equivalently, by fitting the AR(1) on window data and forecasting forward).

5.8 Forecast error metrics

The forecast error at horizon h is denoted as $e_{t+h} = \pi_{t+h} - \widehat{\pi}_{t+h|t}$. In this paper, we use two measures of the error, the root mean squares error and mean absolute error, which are calculated as expressed in the next equations:

$$\mathrm{RMSE}_h = \sqrt{\frac{1}{N_h^{\mathrm{test}}} \sum_{t\in\mathcal{O}} \left(e_{t+h}\right)^2}, \tag{13}$$

$$\mathrm{MAE}_h = \frac{1}{N_h^{\mathrm{test}}} \sum_{t\in\mathcal{O}} \left|e_{t+h}\right|, \tag{14}$$

where $\mathcal{O}$ indexes test forecast origins and N_h^{test} is the number of valid out-of-sample forecasts at horizon h.

5.9 Predictive Ability Evaluation

We implement the Diebold-Mariano test suggested by Diebold and Mariano (1995) to evaluate whether differences in forecast accuracy are statistically significant. Suppose $\ell(\cdot)$ a loss function, the loss differential between model A and benchmark B can be stated as:

$$d_t = \ell(e_{A,t+h}) - \ell(e_{B,t+h})\,. \tag{15}$$

The null hypothesis is equal predictive accuracy,

$$H_0 :\ \mathbb{E}[d_t] = 0. \tag{16}$$

The DM statistic is

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d})}}, \qquad \bar{d} = \frac{1}{N}\sum_{t=1}^{N} d_t, \tag{17}$$

where $\widehat{\mathrm{Var}}(\bar{d})$ is estimated using a heteroskedasticity-and-autocorrelation-consistent (HAC) Newey–West estimator. The HAC variance estimator is:

$$\widehat{S}(q) = \hat{\gamma}_0 + 2\sum_{k=1}^{q} w_k\,\hat{\gamma}_k, \qquad w_k = 1 - \frac{k}{q+1}, \tag{18}$$

with sample autocovariances $\hat{\gamma}_k$. Then

$$\widehat{\mathrm{Var}}(\bar{d}) = \frac{\widehat{S}(q)}{N}. \tag{19}$$

We report DM statistics and two-sided p-values under asymptotic normality. As a sensitivity check, we also consider a smaller lag choice to stabilize inference in shorter samples, following Harvey et al. (1997) and Faust and Wright (2013).

$$q_{\text{sens}} = \min\left(h-1, \left\lfloor (N-1)^{1/3} \right\rfloor\right), \tag{20}$$

In summary, for each horizon $h \in \{1,3,6,12\}$ and each ML method, we implement this procedure: First, we construct horizon-specific feature sets $X_t^{(h)}$ considering publication lags, then split the sample into pre-2024 training data and 2024–2025 test data. The next step is to tune hyper-parameters on the training data using time-series CV and Bayesian optimization, and then we evaluate out-of-sample forecasts using a rolling window of $W = 90$ months, re-estimating the model at each origin. After that, we compute RMSE, MAE, and related summary statistics and compare each model against the benchmarks. Additionally, we estimate the DM test using HAC Newey-West standard errors to check the predictive ability of the models are statistically significant. Table 2 presents a summary of the main characteristics of the methodology implemented in this paper.

Table 2: Experimental design and evaluation protocol (common to all models)

Category	**Specification**
Models	Random Forest (RF), XGBoost (XGB), Ridge, Lasso, Elastic Net (EN), BART, and the hybrid BART+VAR model.
Target and horizons	Inflation (year-over-year CPI inflation (π_t)); $horizons$h $\in \{1,3,6,12\}$ months ahead.
Train / test split	Training: pre-2024. Testing: 2024–2025 (out-of-sample).
OOS evaluation	Rolling re-estimation using a fixed-length window of 90 months (120 months for $h = 12$ when needed).
Tuning and CV	Hyperparameters are selected via Bayesian optimization (`skopt.BayesSearchCV`) on the training sample, separately by horizon. Time-series cross-validation uses rolling/blocked splits that preserve temporal ordering. Objective: minimize RMSE.
Benchmarks	Random walk (RW): $\widehat{\pi}_{t+h\|t} = \pi_t$. AR(1): estimated on the rolling window and iterated h steps ahead.
Accuracy metrics	Test performance is summarized using RMSE and MAE computed over 2024–2025. Loss functions used for model comparison are squared error (SE) and absolute error (AE).
Statistical tests	Diebold–Mariano tests compare each model against RW and AR(1), using Newey–West HAC variance. Baseline truncation lag $q = h-1$; sensitivity lag $q = \min\{h-1, \lfloor (n-1)^{1/3} \rfloor\}$. DM statistics are reported for both SE and AE loss differentials.

Notes: The protocol is held fixed across models; only the estimator and its tuned hyperparameters differ by method and horizon. Time-series cross-validation is applied exclusively within the training sample to prevent information leakage from the test period.

5.10 Machine Learning Methods Implemented

Random Forests. This algorithm is proposed by Breiman (2001). It provides a flexible nonparametric approach to approximate nonlinear relationships by aggregating a large collection of regression trees. Each individual tree recursively partitions the predictor space defined by Z_t into disjoint regions $\{R_1, \ldots, R_M\}$. In its simplest form, a regression tree assigns to each terminal node the average value of the target variable within that region. Hence, the tree-based forecast can be written as

$$\widehat{y}_{t+h} = \sum_{m=1}^{M} c_m \, \mathbf{1}\{Z_t \in R_m\}, \tag{21}$$

where c_m denotes the mean of y_{t+h} in region R_m, and $\mathbf{1}\{\cdot\}$ is the indicator function and Z_t is a vector of covariates. To overcome the instability and high variance of individual regression trees, the algorithm build a large ensemble of trees using bootstrap resamples of the data. At each split, only a random subset of predictors is considered, which reduces correlation across trees and improves generalization performance. The final forecast is obtained by averaging the predictions from all trees in the ensemble.

Extreme Gradient Boosting (XGBoost). Extreme Gradient Boosting (XGBoost), introduced by Chen and Guestrin (2016), is a tree-based ensemble method that builds regression trees sequentially using gradient boosting, detailed on Friedman (2001). Unlike Random Forests, which aggregate independently grown trees, XGBoost constructs trees in an additive manner, where each new tree is fitted to the residuals of the previous ensemble in order to minimize a regularized objective function.

Formally, the h-step-ahead forecast is represented as

$$\widehat{y}_{t+h} = \sum_{k=1}^{K} f_k(Z_t), \tag{22}$$

where each $f_k(\cdot)$ is a regression tree belonging to a predefined function space $\mathcal{F}$, and Z_t is the vector of covariates. Each tree partitions the predictor space into disjoint regions and assigns a constant prediction to each terminal node, similarly to standard regression trees. However, trees are added sequentially to minimize the regularized objective

$$\mathcal{L} = \sum_{t=1}^{T} \ell(y_{t+h}, \widehat{y}_{t+h}) + \sum_{k=1}^{K} \Omega(f_k), \tag{23}$$

where $\ell(\cdot)$ is a differentiable loss function (in our case, squared error) and $\Omega(f_k)$ is a regularization term that penalizes model complexity, typically depending on the number of leaves and the magnitude of leaf weights.

At each boosting iteration, the algorithm performs a second-order Taylor expansion of the loss function to approximate the optimization problem, which allows efficient computation of optimal splits and leaf weights. Regularization parameters, shrinkage (learning rate), subsampling of observations, and column subsampling are used to prevent overfitting and improve out-of-sample performance.

The final forecast is obtained by adding the contributions of all K trees. Hyperparameters governing tree depth, number of boosting rounds, learning rate, subsampling rates, and regularization terms are selected via Bayesian optimization using time-series cross-validation, as described previously.

Elastic Net. Proposed by Zou and Hastie (2005), it combines Lasso and Ridge models. Particularly, in the Elastic Net the Lasso estimator corresponds with the case when ($\alpha = 1$), while the if ($\alpha = 0$) we get the Ridge estimator. The Elastic Net estimator solves the penalized least-squares problem

$$\widehat{\beta} = \arg\min_{\beta} \left\{ \sum_{t=1}^{T} \left(y_{t+h} - Z_t\beta\right)^2 + \lambda \sum_{i=1}^{N_Z} \left(\alpha|\beta_i| + (1-\alpha)\beta_i^2\right) \right\}, \tag{24}$$

where $\beta = (\beta_1, \ldots, \beta_{N_Z})'$ is the parameter vector, while $\lambda \geq 0$ and $\alpha \in [0, 1]$ are hyperparameters.

Bayesian Additive Regression Trees (BART). Introduced by Chipman et al. (2010), is a Bayesian nonparametric ensemble method that represents the regression function as a sum of many weak regression trees. Unlike boosting algorithms such as XGBoost, which estimate trees sequentially through gradient-based optimization, BART estimates all trees jointly within a fully probabilistic framework using Bayesian back-fitting. Formally, the h-step-ahead forecast is modeled as

$$y_{t+h} = \sum_{k=1}^{K} g_k(Z_t; \mathcal{T}_k, \mathcal{M}_k) + \varepsilon_{t+h}, \qquad \varepsilon_{t+h} \sim \mathcal{N}(0, \sigma^2), \tag{25}$$

where each $g_k(\cdot)$ denotes a regression tree characterized by a tree structure $\mathcal{T}_k$ and a set of terminal node parameters $\mathcal{M}_k$. The predictor vector Z_t is defined as in the previous specifications. Each tree partitions the predictor space into disjoint regions and assigns a constant value to each terminal node. The overall regression function is the sum of K such trees.

BART imposes a set of regularizing priors designed to prevent overfitting while retaining substantial flexibility. Each regression tree is assigned a prior on its structure such that a node at depth d splits with probability

$$\Pr(\text{split at depth } d) = \alpha(1+d)^{-\beta}, \tag{26}$$

with typical default values $\alpha = 0.95$ and $\beta = 2$. This formulation strongly favors shallow trees, ensuring that individual trees act as weak learners.

Conditional on a given tree structure, terminal node parameters are assigned Gaussian priors,

$$\mu_{kj} \sim \mathcal{N}(0, \sigma_\mu^2), \tag{27}$$

where σ_μ is scaled inversely with the number of trees K, inducing shrinkage at the tree level. The error variance follows an inverse-gamma (or inverse-chi-square) prior,

$$\sigma^2 \sim \text{IG}\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right), \tag{28}$$

with hyperparameters chosen to provide weakly informative but stabilizing regularization.

Together, these priors ensure that flexibility arises from the aggregation of many small trees rather than from overly complex individual components, yielding a robust and regularized nonparametric regression framework.

Regarding the posterior distribution, inference in BART is conducted within a fully Bayesian framework. Given the likelihood

$$y_i = \sum_{k=1}^{K} g_k(x_i; \mathcal{T}_k, \mathcal{M}_k) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{29}$$

and the prior specification on tree structures, terminal node parameters, and the error variance, posterior inference proceeds from

$$p\big(\{\mathcal{T}_k, \mathcal{M}_k\}_{k=1}^{K}, \sigma^2 \mid \mathbf{y}\big) \propto p(\mathbf{y} \mid \{\mathcal{T}_k, \mathcal{M}_k\}, \sigma^2) \prod_{k=1}^{K} p(\mathcal{T}_k) p(\mathcal{M}_k \mid \mathcal{T}_k) p(\sigma^2). \tag{30}$$

In practice, this posterior is approximated via Markov Chain Monte Carlo (MCMC), typically using a Gibbs sampler with Bayesian backfitting. At each iteration, trees are updated sequentially conditional on the partial residuals obtained from all other trees, and σ^2 is sampled from its full conditional distribution.

The posterior distribution over the regression function naturally yields point forecasts given by the posterior mean,

$$\widehat{y}_{t+h} = \mathbb{E}[y_{t+h} \mid Z_t], \tag{31}$$

as well as predictive uncertainty via posterior credible intervals. This probabilistic characterization of the forecasting function distinguishes BART from purely frequentist ensemble methods and provides coherent uncertainty quantification in out-of-sample prediction.

Unlike the remaining machine learning models, BART (and VAR+BART method which is presented later in this section) were estimated using the default prior specification proposed by Chipman et al. (2010). This choice reflects both computational considerations and the extensive evidence in the literature showing that BART tends to perform robustly under its default regularization settings. Consequently, BART results should be interpreted as conservative estimates of its forecasting performance, as further gains could potentially be obtained through extensive hyperparameter tuning.

BART+VAR residual-hybrid model. Following (Zhang, 2003) and (de O. Santos Júnior et al., 2023), we can think on a time series as a sum of a linear and a nonlinear components, and then implement a residual-hybrid forecasting model that combines a linear vector autoregressive (VAR) specification with a nonlinear correction based on Bayesian Additive Regression Trees (BART). If we let the VAR(p) model deliver an h-step-ahead forecast $\widehat{\pi}^{VAR}_{t+h|t}$ for inflation, then the forecast error of the linear block is

$$e_t^{(h)} = \pi_{t+h} - \widehat{\pi}^{VAR}_{t+h|t}. \tag{32}$$

Later, a BART can be used to approximate the conditional expectation of this residual given a potentially large information set Z_t,

$$e_t^{(h)} = f_h(Z_t) + \varepsilon_t^{(h)}, \tag{33}$$

where $f_h(\cdot)$ is modeled as a sum of regression trees under the Bayesian additive regression trees framework of Chipman et al. (2010). The final hybrid forecast is obtained additively as

$$\widehat{\pi}^{HYB}_{t+h|t} = \widehat{\pi}^{VAR}_{t+h|t} + \widehat{e}^{BART}_{t+h|t}. \tag{34}$$

This approach is relatively uncommon in macroeconomic forecasting, however it is interesting as it combines a traditional method with an ML technique that might improve its forecast performance.

Conceptually, the residual-hybrid structure preserves the interpretability and stability of the linear VAR backbone while allowing BART to flexibly capture nonlinearities, interactions, and threshold effects that may be present in the inflation process but are not well approximated by linear dynamics alone. A similar approach for combining a VAR model and a BART model is BAVART developed by Huber and Rossini (2021), nevertheless, this approach is not developed in this paper.

VAR block. We estimate a reduced-form vector autoregression (VAR) with $p = 6$ lags on a small set of macro-financial variables selected to capture external cost pressures, domestic prices, expectations, and the monetary policy stance. Let's denote

$$X_t = \big(\Delta \mathrm{WTI}_t,\ \mathrm{PPI}^{\mathrm{man}}_t,\ \Delta \mathrm{NER}_t,\ \Delta \pi^{e,12m}_t,\ \pi_t,\ i_t\big)',$$

where π_t denotes year-over-year inflation, i_t is the monetary policy rate, $\Delta \mathrm{WTI}_t$ is the first difference of the WTI oil price, $\mathrm{PPI}^{\mathrm{man}}_t$ is the manufacturing producer price index in levels, $\Delta \mathrm{NER}_t$ is the first difference of the nominal exchange rate, and $\Delta \pi^{e,12m}_t$ is the first difference of 12-month-ahead inflation expectations (survey-based). The VAR is then

$$X_t = c + \sum_{j=1}^{p} A_j X_{t-j} + u_t, \qquad p = 6, \tag{35}$$

with u_t a mean-zero innovation vector and $(A_j)_{j=1}^{p}$ coefficient matrices.

Note that some variables were transformed to maintain the stationarity requirement for a VAR estimation. We kept inflation (π_t), the policy rate (i_t), and the manufacturing PPI in levels, while WTI, the nominal exchange rate, and inflation expectations enter in first differences (as indicated in X_t). The model is estimated by OLS equation-by-equation on each estimation window [2].

The VAR block is not intended to provide structural inference. Rather, it serves as a parsimonious linear forecasting component within the hybrid framework. The objective is to capture the main linear comovements among a small set of economically relevant variables, while nonlinear patterns are subsequently modeled through BART.

6 Results

The main out-of-sample (OOS) forecasting results for year-over-year inflation over the 2024–2025 evaluation period are presented in Table 3. Forecast accuracy is assessed using RMSE and MAE across four horizons, $h \in \{1, 3, 6, 12\}$ months ahead. All models are estimated under a common rolling-window design, with hyperparameters selected via Bayesian optimization within the training sample and evaluated against two standard benchmarks: the Random Walk (RW) and an AR(1) model[3]

However, it should be mentioned that while this design allows us to assess forecasting performance under the most recent inflation regime, the effective number of observations decreases substantially as the forecast horizon increases. In particular, the h = 12 evaluation is based on only twelve observations. Therefore, results and statistical inference at long horizons should be interpreted as exploratory and indicative rather than definitive.

Table 3: Main out-of-sample forecasting performance (2024–2025)

	H+1		H+3		H+6		H+12	
Model	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Random Forest	0.6035	0.5158	0.7694	0.6472	0.9636	0.7856	1.6782	1.4509
XGBoost	0.5550	0.4789	**0.6204**	**0.5081**	0.8458	0.6850	0.7781	0.6457
BART	0.5208	0.4008	0.6334	0.5247	**0.7403**	**0.6312**	0.6789	0.5757
BART+VAR	**0.4727**	**0.3861**	0.8584	0.7441	0.9192	0.7155	1.1046	1.0164
Ridge	0.6169	0.5271	0.9460	0.7804	1.1597	1.0103	0.7717	0.6334
Lasso	0.5393	0.4337	0.9866	0.8065	1.2496	1.0422	**0.6193**	**0.5191**
Elastic Net	0.5396	0.4335	0.9859	0.8058	1.2486	1.0441	0.8343	0.6635
RW	0.7604	0.6552	1.1064	0.9557	1.3115	1.1567	1.6037	1.2583
AR(1)	0.7475	0.6370	1.0909	0.9435	1.2952	1.1336	1.5780	1.2415

Notes: Entries report out-of-sample RMSE and MAE over 2024–2025 for each forecasting horizon. RW denotes the random-walk benchmark and AR(1) the autoregressive benchmark.

"Nowcasting" (t+1). At the one-month horizon, nonlinear machine learning methods clearly dominate both linear specifications and benchmark models. The hybrid **BART+VAR** model delivers the best performance (RMSE = 0.4727, MAE = 0.3861), followed by **BART** (RMSE = 0.5208) and **XGBoost** (RMSE = 0.5550).

[2] For completeness, a recursive (Cholesky) ordering in levels given by $(\mathrm{WTI}, \mathrm{PPI}^{\mathrm{man}}, \mathrm{NER}, \pi^{e,12m}, \pi, i)$ can be used to define a contemporaneous structure for shock identification within the VAR. However, the hybrid forecasting exercise relies exclusively on reduced-form forecasts and residuals. Consequently, this identification scheme is not used in forecast generation and has no effect on the forecasting results reported in this paper.

[3] To complement this section, annex 9.1 includes an analysis of the SHAP values for the XGB model, as an attempt to present some economic interpretability to the results obtained.

The improvement relative to the Random Walk (RMSE = 0.7604) and AR(1) (RMSE = 0.7475) is substantial, which suggests that short-run inflation dynamics are not well approximated by simple persistence models. The significant performance of tree-based nonlinear estimators over the rest of the models suggests that inflation at very short horizons contains nonlinear interactions and threshold effects that linear shrinkage methods (Ridge, Lasso, Elastic Net) only partially capture, as discussed in the Literature Review section. Additionally, the results of the hybrid specification are consistent with the idea that the VAR block captures structured macroeconomic co-movements, while BART flexibly models residual nonlinearities.

Short term (t+3) and (t+6). At the three-month horizon, **XGBoost** emerges as the best-performing model (RMSE = 0.6204), followed closely by **BART** (RMSE = 0.6334). The hybrid BART+VAR model performs less favorably at this horizon (RMSE = 0.8584), suggesting that the additive boosting structure of XGBoost may better capture medium-run nonlinear patterns.

At six months ahead, **BART** outperforms all competitors (RMSE = 0.7403, MAE = 0.6312). Again, the nonlinear ensemble models maintain a consistent advantage over linear regularized methods, whose RMSE values exceed 1.00 in some cases, which reinforces the idea that inflation dynamics become increasingly nonlinear at these horizons.

Long horizon (t+12). At the twelve-month horizon, the ranking shifts. **Lasso** delivers the best performance (RMSE = 0.6193, MAE = 0.5191), followed by **Ridge** and **XGBoost**. In contrast, the hybrid BART+VAR model deteriorates (RMSE = 1.1046).

One interpretation of this result is that while nonlinear residual corrections improve short-run performance, longer-horizon forecasts may benefit from stronger shrinkage and more parsimonious structures.Additionally, note that all multivariate models outperform the Random Walk and AR(1) benchmarks at t+12, confirming that macroeconomic information improves predictive accuracy even at longer horizons.

In short, the results suggest that i) most multivariate models outperform the benchmark specifications across forecasting horizons, although some methods, particularly Random Forest at h = 12, do not consistently dominate the benchmark models; ii) the nonlinear ensemble methods dominate at short horizons (up to h=6), suggesting the presence of nonlinearities that are relevant in the inflation dynamics, and iii) the linear models regain competitiveness at longer horizons, where overfitting risks and parameter instability might increase.

Economic Interpretation and Policy Implications. Taken together, these results suggest that inflation dynamics in Costa Rica have become increasingly state-dependent over the sample period. During much of the pre-pandemic period, particularly after the adoption of inflation targeting in 2018, inflation remained relatively low and stable, with well-anchored expectations and limited volatility in domestic macroeconomic conditions. In such an environment, the relationships between inflation and its determinants are likely to be more stable and adequately approximated by linear models.

However, the period following 2020 was characterized by an unusual succession of shocks, including the COVID-19 pandemic, global supply-chain disruptions, sharp increases in commodity and transportation costs, rapid monetary policy tightening across advanced economies, and subsequent disinflationary pressures influenced by lower international prices and exchange-rate appreciation. These developments likely altered the transmission mechanisms of inflation and generated structural changes that are difficult to capture using fixed linear relationships. The superior performance of nonlinear methods at short horizons is consistent with the presence of these regime-dependent dynamics.

Several forms of nonlinearity may be particularly relevant in the Costa Rican case. First, the pass-through from international commodity prices and exchange-rate movements to domestic inflation is unlikely to be constant over time. Large increases in oil prices and imported input costs, such as those observed during 2021–2022, may generate stronger and faster inflationary responses than equivalent declines, reflecting pricing rigidities, adjustment costs, or asymmetric transmission mechanisms. Second, inflation expectations may become more informative when they move sufficiently away from the Central Bank's target, generating threshold effects that are difficult to capture through linear specifications. Third, the interaction between economic activity, labor-market conditions, and monetary policy may vary across different phases of the business cycle, particularly during periods of heightened uncertainty. Tree-based methods such as XGBoost and BART are well suited to detect these types of nonlinear interactions and regime changes, which is consistent with their superior forecasting performance at short horizons reported in Table 3. In contrast, the stronger performance of penalized linear models at longer horizons suggests that medium-term inflation dynamics remain anchored by more persistent and stable macroeconomic relationships, where parsimonious specifications may offer greater robustness against overfitting.

Additional evidence supporting this interpretation is provided by the SHAP analysis presented in Annex 9.1. The results reveal a clear horizon-dependent structure in inflation forecasting. At short horizons, inflation persistence, producer prices, wages, and inflation expectations emerge among the most influential predictors, suggesting that inertia and cost pressures are key drivers of near-term inflation dynamics. As the forecasting horizon expands, variables associated with external trade, exchange-rate movements, and financial conditions gain relevance, while the monetary policy rate becomes the most important predictor at the annual horizon. This pattern is consistent with the view that short-run inflation fluctuations are largely driven by cost and external shocks, whereas medium- and long-term inflation dynamics remain anchored by monetary policy and broader macroeconomic fundamentals.

From a policy perspective, these findings suggest that central banks may benefit from combining nonlinear machine learning tools for short-term projections and shrinkage-based approaches for longer-term projections. The superior performance of nonlinear methods at horizons up to six months indicates that these models can provide valuable information for assessing near-term inflationary pressures, identifying emerging risks, and supporting near-term monetary policy assessment in environments characterized by elevated uncertainty or rapidly changing economic conditions. At the same time, the stronger performance of more parsimonious linear specifications at longer horizons suggests that medium-term inflation dynamics remain closely linked to relatively stable macroeconomic fundamentals and well-anchored expectations.

Consequently, rather than viewing machine learning methods as substitutes for traditional forecasting approaches, the results support a complementary framework in which different models provide complementary information according to their comparative advantages across forecasting horizons. For a small and highly open economy such as Costa Rica, where inflation is strongly influenced by external shocks and changing global conditions, maintaining a diversified forecasting toolkit may be particularly valuable for supporting monetary policy decisions under uncertainty.

Table 4: Diebold–Mariano tests (out-of-sample): models vs. benchmarks (squared error loss).

		vs. Random Walk (RW)				vs. AR(1)			
Horizon	Model	DM	p	n	Sig.	DM	p	n	Sig.
H+1	Random Forest	2.0632	0.0391	23	**	2.2595	0.0239	23	**
H+1	XGBoost	-2.7362	0.0062	23	***	-2.5753	0.0100	23	**
H+1	BART	-2.6488	0.0081	23	***	-2.5907	0.0096	23	***
H+1	BART+VAR	0.0131	0.9895	22		0.1926	0.8473	22	
H+1	Ridge	-2.5278	0.0115	23	**	-2.1886	0.0286	23	**
H+1	Lasso	-2.4601	0.0139	23	**	-2.3694	0.0178	23	**
H+1	Elastic Net	-2.4569	0.0140	23	**	-2.3655	0.0180	23	**
H+3	Random Forest	-1.2869	0.1981	21		-1.4567	0.1452	21	
H+3	XGBoost	-3.0335	0.0024	21	***	-3.0784	0.0021	21	***
H+3	BART	-2.5367	0.0112	21	**	-2.4965	0.0125	21	**
H+3	BART+VAR	-1.1436	0.2528	21		-1.4076	0.1592	21	
H+3	Ridge	-0.9274	0.3537	21		-0.8472	0.3969	21	
H+3	Lasso	-0.8160	0.4145	21		-0.7694	0.4417	21	
H+3	Elastic Net	-0.8213	0.4115	21		-0.7751	0.4383	21	
H+6	Random Forest	-2.5602	0.0105	18	**	-2.4683	0.0136	18	**
H+6	XGBoost	-3.4390	0.0006	18	***	-3.4857	0.0005	18	***
H+6	BART	-2.9429	0.0033	18	***	-2.7250	0.0064	18	***
H+6	BART+VAR	-2.1948	0.0282	18	**	-1.3075	0.1910	18	
H+6	Ridge	-0.8377	0.4022	18		-0.8097	0.4181	18	
H+6	Lasso	-0.2990	0.7649	18		-0.2366	0.8129	18	
H+6	Elastic Net	-0.3024	0.7624	18		-0.2406	0.8099	18	
H+12	Random Forest	*see note*	–	12		*see note*	–	12	
H+12	XGBoost	-1.5538	0.1202	12		-1.5795	0.1142	12	
H+12	BART	-1.6497	0.0990	12	*	-1.6631	0.0963	12	*
H+12	BART+VAR	-1.5835	0.1133	12		-1.6630	0.0963	12	*
H+12	Ridge	-1.6025	0.1091	12		-1.6326	0.1026	12	
H+12	Lasso	-1.7925	0.0730	12	*	-1.8353	0.0665	12	*
H+12	Elastic Net	-1.5446	0.1224	12		-1.5690	0.1167	12	

Source: Authors' calculations.
Notes: Entries report Diebold–Mariano statistics using squared error loss (SE) and HAC Newey–West variance with baseline lag $q = h - 1$. Negative DM statistics indicate that the model outperforms the benchmark (lower loss). Significance codes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.
Note on Random Forest at $h = 12$: some baseline DM statistics are numerically unstable (extreme magnitudes / missing values) in the raw output; the corresponding inference should rely on the sensitivity specification and be interpreted cautiously.

Evidence of predictive ability. Table 4 summarizes Diebold–Mariano (DM) tests comparing each model against two standard benchmarks, a random walk (RW) and an AR(1). Overall, the strongest and most consistent rejections of equal predictive accuracy occur at short horizons (up to 6 months) for the nonlinear tree-based methods, particularly XGBoost and BART. At $h = 1$, both XGBoost and BART deliver statistically significant improvements over RW and AR(1), while the linear models also show gains at the one-month horizon, but less significant.

At $h = 3$, XGBoost remains the dominant model, rejecting equal accuracy against both benchmarks at the 1% level of significance, followed by the BART model which is also superior to both benchmarks. In contrast, the linear models are no longer statistically significant at this horizon.

At $h = 6$, the ranking is broadly similar: XGBoost and BART and Random Forest continue to outperform both benchmarks with strong statistical support. The hybrid BART+VAR specification shows some evidence of improvement versus RW at $h = 6$, but the advantage is not robust against AR(1) within the baseline specification.

At $h = 12$, statistical power declines influenced by the shorter effective evaluation sample, and most methods fail to reject equal accuracy at the 5% level of significance. Still, BART and Lasso provide suggestive evidence of improvements against the benchmarks, whereas XGBoost loses its forecast performance at the annual horizon in this evaluation. These results are consistent with the view that short-horizon inflation forecasting benefits from flexible nonlinear approximations, while longer-horizon performance is harder to distinguish statistically given limited out-of-sample observations.

In terms of performance between linear and nonlinear models, the DM results support the hypothesis that nonlinear interactions and threshold-like effects in macroeconomic predictors are relevant for near-term inflation dynamics. Tables A3 to A9 present the detailed Diebold–Mariano test results, including the sensitivity checks, which are also consistent with the main findings.

7 Conclusions

In this paper, we estimated machine learning models to produce inflation forecasts for the Costa Rican economy at various horizons, with a strategy that used a common rolling-window design and Bayesian hyperparameter selection. Our results show that multivariate information yields significant gains over naive persistence benchmarks. Across horizons, most models improve upon the Random Walk and AR(1), confirming that macroeconomic covariates contain predictive content beyond simple inflation inertia. This finding is economically relevant, as it suggests that there are signals in the information set that can be exploited to improve forecasts.

The forecast performance of the models suggests that for short horizons, tree-based ensemble methods dominate: BART+VAR is best at $h = 1$, XGBoost leads at $h = 3$, and BART performs best at $h = 6$. Besides, the Diebold–Mariano tests reinforce this ranking, with the most consistent and statistically significant rejections of equal predictive accuracy at short horizons for XGBoost and BART. The evidence is consistent with the hypothesis that near-term inflation dynamics is influenced by nonlinearities and interaction effects that are difficult to capture by linear shrinkage specifications.

At the twelve-month horizon, the ranking shifts and more parsimonious linear models regain competitiveness, with Lasso delivering the best point-forecast accuracy and providing suggestive DM evidence against benchmarks.

However, a limitation of this study is the relatively short out-of-sample evaluation period (2024–2025). As the forecast horizon increases, the effective number of evaluation observations decreases substantially, reaching only twelve observations for the twelve-month-ahead forecasts. This limits the statistical power of forecast comparison tests and may affect the stability of some Diebold-Mariano statistics. Therefore, results at longer horizons should be interpreted with caution, and future research could reassess these findings as additional data become available.

From a policy perspective, the results support a pragmatic forecasting strategy that combines nonlinear machine learning tools for near-term projections with shrinkage-based linear approaches for longer-horizon guidance. Rather than viewing these methods as competing alternatives, the evidence suggests that they provide complementary information across forecasting horizons. For central banks operating in small and highly open economies such as Costa Rica, maintaining a diversified forecasting toolkit may improve forecast robustness, strengthen the assessment of inflationary risks, and enhance the information available for monetary policy decision-making.

8 Bibliography

Abraham, K. G., Jarmin, R. S., Moyer, B., and Shapiro, M. D. (2022). *Big Data for Twenty-First-Century Economic Statistics*. University of Chicago Press.

Araujo, G. S. and Gaglianone, W. P. (2023). Machine learning methods for inflation forecasting in brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2):100087.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pages 785–794, New York, NY, USA. ACM.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266 – 298.

Cicceri, G., Inserra, G., and Limosani, M. (2020). A machine learning approach to forecast economic recessions—an italian case study. *Mathematics*.

de O. Santos Júnior, D. S., de Mattos Neto, P. S., de Oliveira, J. F., and Cavalcanti, G. D. (2023). A hybrid system based on ensemble learning to model residuals for time series forecasting. *Information Sciences*, 649:119614.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Faust, J. and Wright, J. H. (2013). Chapter 1 - forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 2–56. Elsevier.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Goulet Coulombe, P., Marcellino, M., and Stevanović, D. (2021). Can machine learning catch the covid-19 recession? *National Institute Economic Review*, 256:71–109.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., and Schreiner, J. (2023). Nowcasting in a pandemic using non-parametric mixed frequency vars. *Journal of Econometrics*, 232(1):52–69.

Huber, F. and Rossini, L. (2021). Inference in bayesian additive vector autoregressive tree models.

Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? a case study of u.s. consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Liu, Y. (2024). Mending the Crystal Ball: Enhanced Inflation Forecasts with Machine Learning. *IMF Working Papers*, 2024(206):1.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions.

Magazzino, C., Mele, M., and Mutascu, M. (2025). An artificial neural network experiment on the prediction of the unemployment rate. *Journal of Policy Modeling*, 47(3):471–491.

Medeiros, M. C. and Vasconcelos, G. F. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, 138:50–52.

Medeiros, M. C., Vasconcelos, G. F. R., Álvaro Veiga, and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119.

Nyman, R. and Ormerod, P. (2017). Predicting economic recessions using machine learning algorithms. *arXiv: General Finance*.

Rodríguez-Vargas, A. (2022). Estimation of potential output for costa rica. 1995-2021. Technical report, Banco Central de Costa Rica.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Neural Information Processing Systems*.

Tehranian, K. (2023). Can machine learning catch economic recessions using economic and market sentiments? *SSRN Electronic Journal*.

Zhang, G. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

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.

9 Annex

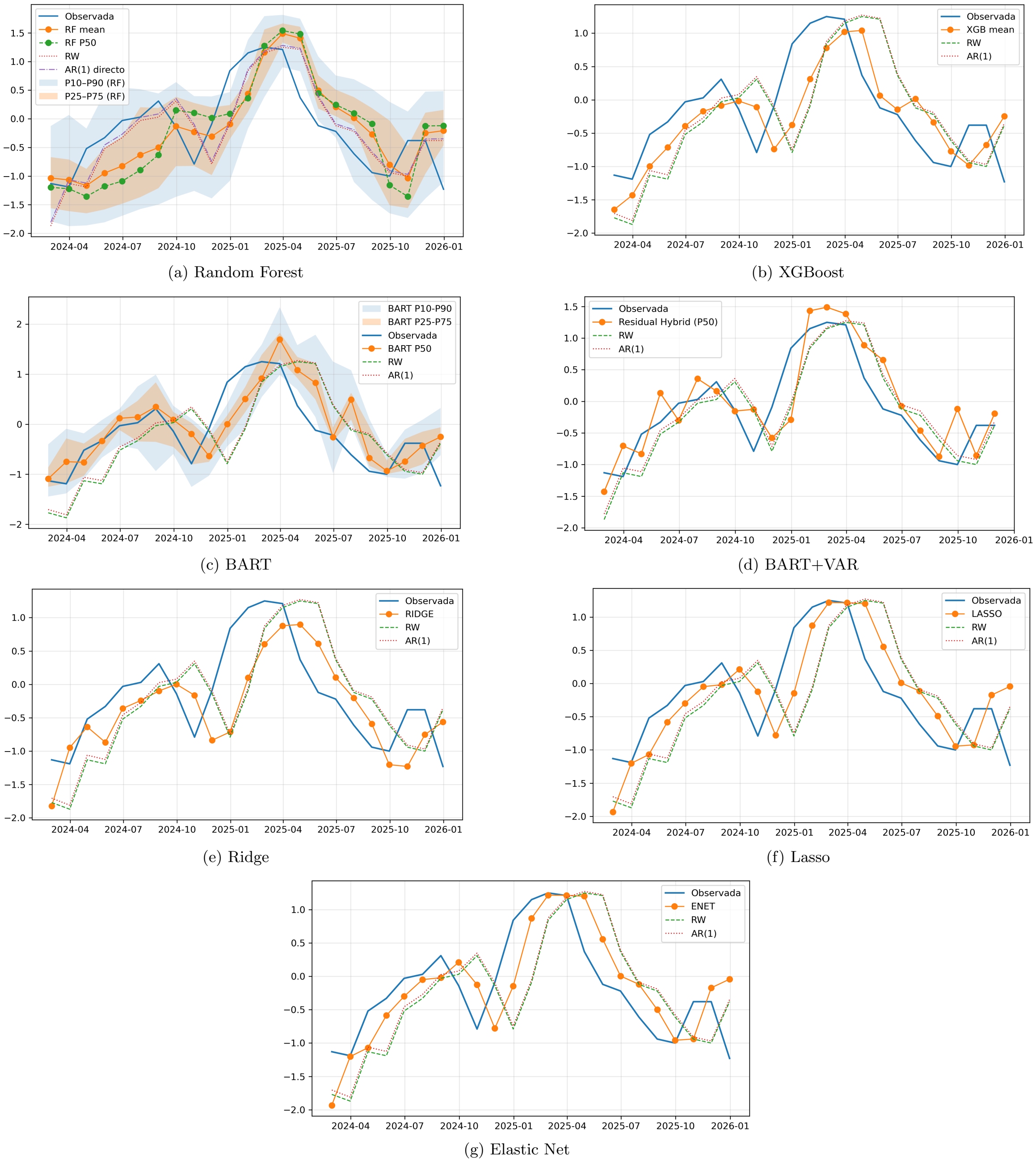
Table A1: Descriptive statistics of the variables in the dataset

	N	Mean	Std.Dev.	Min	P10	Median	P90	Max
inf_int	168	2.276	2.784	-3.280	-0.866	1.835	5.712	12.130
ipp_man	168	2.682	4.777	-5.720	-2.522	2.260	9.470	17.590
ip_mpi	168	1.995	29.806	-45.338	-26.668	-4.386	33.512	139.234
wti	168	3.338	38.655	-71.468	-37.261	-3.714	53.019	238.664
ismn	168	3.589	2.274	0.484	1.149	2.598	6.814	8.139
ismr	168	1.334	2.787	-7.655	-1.333	1.470	3.774	10.434
w_nom_tc	167	2.506	4.363	-8.070	-3.587	3.650	7.289	9.616
w_real_tc	167	0.276	5.057	-10.014	-6.859	0.966	6.539	9.397
imae	167	3.526	3.650	-8.980	1.362	3.780	6.580	13.090
imae_re	167	10.851	6.024	-3.650	3.204	11.140	17.478	30.150
imae_rd	167	2.561	3.754	-9.960	-0.256	2.930	6.024	11.410
imae_sd	167	3.714	6.369	-32.140	-1.278	4.020	9.750	20.290
exp_rd	168	2.668	5.624	-8.300	-4.400	2.900	7.930	21.500
exp_re	168	12.155	8.377	-11.400	2.110	11.400	22.150	39.300
exp_tot	168	7.836	7.008	-9.900	0.110	6.750	15.990	32.900
imp_rd	168	6.221	14.238	-16.100	-10.201	5.127	27.980	47.744
imp_re	168	11.050	16.541	-13.623	-10.179	12.092	36.196	50.526
imp_tot	168	7.042	14.372	-16.300	-9.600	5.150	30.981	48.400
cred_mn	167	8.535	4.908	0.661	2.878	7.120	15.318	20.090
cred_me	167	6.448	7.482	-5.429	-3.436	7.392	16.023	19.634
cred_tot	167	7.733	4.895	-0.106	0.376	7.104	14.241	16.039
exp_inf_12_prom	168	4.146	1.624	1.300	2.049	3.664	6.347	9.893
exp_inf_12_med	168	3.990	1.544	1.000	2.000	3.500	6.000	9.500
exp_inf_12_merc	168	2.657	0.885	0.648	1.576	2.572	3.656	4.967
brecha_exp_enc	168	1.229	1.098	-0.300	0.108	0.905	2.552	6.890
brecha_exp_mer	168	-0.954	0.953	-3.032	-2.218	-1.006	0.444	1.468
exp_inf_12_enc_var	168	-0.359	1.615	-6.748	-2.003	-0.264	0.592	6.936
exp_inf_12_merc_var	168	-0.182	1.163	-2.801	-1.681	-0.228	1.357	2.805
tnp	167	59.835	2.684	53.192	55.834	60.118	62.897	63.861
to	167	53.341	2.819	43.135	49.928	53.690	56.400	57.794
ti	167	40.165	2.684	36.139	37.103	39.882	44.166	46.808
td	167	10.816	3.555	5.744	7.449	9.726	15.446	24.434
ts	167	10.175	4.520	2.757	3.668	10.000	14.282	26.191
res_fin_gc	166	-4.798	1.353	-8.566	-6.348	-4.916	-3.118	-2.121
deuda_tot	166	49.304	11.972	26.879	33.776	50.491	63.912	68.262
deuda_int	166	38.046	8.367	23.124	27.246	40.758	47.575	52.061
tbp	168	5.624	1.704	2.800	3.435	5.850	7.200	10.500
tpm	168	4.000	1.977	0.750	0.932	4.365	5.978	9.000
prime	168	4.564	2.045	2.100	2.310	3.570	8.066	8.500
tcn_var	168	0.179	6.293	-21.070	-6.467	0.305	7.091	11.590
basem_var	167	8.252	5.474	-6.830	1.596	8.690	14.650	20.030
m1_var	167	9.711	14.743	-13.020	-5.684	8.180	21.942	71.540
itcer_var	167	-0.694	6.392	-18.090	-8.074	-0.660	6.602	15.320

Table A2: Selected hyperparameters (Bayesian optimization) by horizon and model.

Model	**Horizon**	**Selected hyperparameters**
Random Forest	H+1	bootstrap=True; criterion=squared_error; max_depth=14; max_features=0.85; max_samples=0.893864; min_samples_leaf=7; min_samples_split=8; n_estimators=300
	H+3	bootstrap=True; criterion=squared_error; max_depth=13; max_features=0.25; max_samples=1; min_samples_leaf=3; min_samples_split=6; n_estimators=300
	H+6	bootstrap=True; criterion=squared_error; max_depth=18; max_features=0.95; max_samples=1; min_samples_leaf=3; min_samples_split=5; n_estimators=1423
	H+12	bootstrap=True; criterion=squared_error; max_depth=7; max_features=0.3; max_samples=0.65; min_samples_leaf=20; min_samples_split=5; n_estimators=1196
XGBoost	H+1	colsample_bytree=0.784842; gamma=0.326209; learning_rate=0.0230632; max_depth=2; min_child_weight=10.6561; n_estimators=1717; reg_alpha=0.003; reg_lambda=2; subsample=0.95
	H+3	colsample_bytree=0.853707; gamma=0.00793049; learning_rate=0.0772529; max_depth=6; min_child_weight=5.98357; n_estimators=1443; reg_alpha=7.61735e-06; reg_lambda=17.7949; subsample=0.949718
	H+6	colsample_bytree=0.899463; gamma=0; learning_rate=0.07; max_depth=6; min_child_weight=5; n_estimators=1255; reg_alpha=6.96734e-06; reg_lambda=2.57399; subsample=0.98
	H+12	colsample_bytree=0.8; gamma=0.8; learning_rate=0.06; max_depth=5; min_child_weight=6; n_estimators=900; reg_alpha=0.002; reg_lambda=4.19433; subsample=0.98
BART	H+1	num_trees=200; alpha=0.95; beta=2; num_gfr=10; num_mcmc=300; num_burnin=150; fixed_window=90; fan_ensemble=10; CV_RMSE_pre2024=2.51529
	H+3	num_trees=200; alpha=0.95; beta=2; num_gfr=10; num_mcmc=300; num_burnin=150; fixed_window=90; fan_ensemble=10; CV_RMSE_pre2024=3.39386
	H+6	num_trees=200; alpha=0.95; beta=2; num_gfr=10; num_mcmc=300; num_burnin=150; fixed_window=90; fan_ensemble=10; CV_RMSE_pre2024=3.92193
	H+12	num_trees=300; alpha=0.9; beta=2; num_gfr=10; num_mcmc=300; num_burnin=150; fixed_window=120; fan_ensemble=10; CV_RMSE_pre2024=4.58363
BART+VAR	H+1	num_trees=200; alpha=0.95; beta=2; fixed_window=90; fan_ensemble=50; VAR_spec=Δwti, ipp_man, Δtcn_var, Δexp_inf_12_prom, inf_int, tpm; VAR_lags=6
	H+3	num_trees=200; alpha=0.95; beta=2; fixed_window=90; fan_ensemble=50; VAR_spec=Δwti, ipp_man, Δtcn_var, Δexp_inf_12_prom, inf_int, tpm; VAR_lags=6
	H+6	num_trees=200; alpha=0.95; beta=2; fixed_window=90; fan_ensemble=50; VAR_spec=Δwti, ipp_man, Δtcn_var, Δexp_inf_12_prom, inf_int, tpm; VAR_lags=6
	H+12	num_trees=300; alpha=0.9; fixed_window=120; fan_ensemble=50; VAR_lags=6
Ridge	H+1	ridge__alpha=52.0368
	H+3	ridge__alpha=16.667412
	H+6	ridge__alpha=100.000000
	H+12	ridge__alpha=0.082077
Lasso	H+1	lasso__alpha=0.0965617
	H+3	lasso__alpha=0.083252
	H+6	lasso__alpha=0.056082
	H+12	lasso__alpha=0.000169
Elastic Net	H+1	enet__alpha=0.0967456; enet__l1_ratio=0.99
	H+3	enet__alpha=0.084851; enet__l1_ratio=0.99
	H+6	enet__alpha=0.056225; enet__l1_ratio=0.99
	H+12	enet__alpha=0.010000; enet__l1_ratio=0.30

Source: Authors' calculations.
Notes: Hyperparameters were selected using Bayesian optimization with time-series cross-validation on the pre-2024 sample.



(a) Random Forest

(b) XGBoost

(c) BART

(d) BART+VAR

(e) Ridge

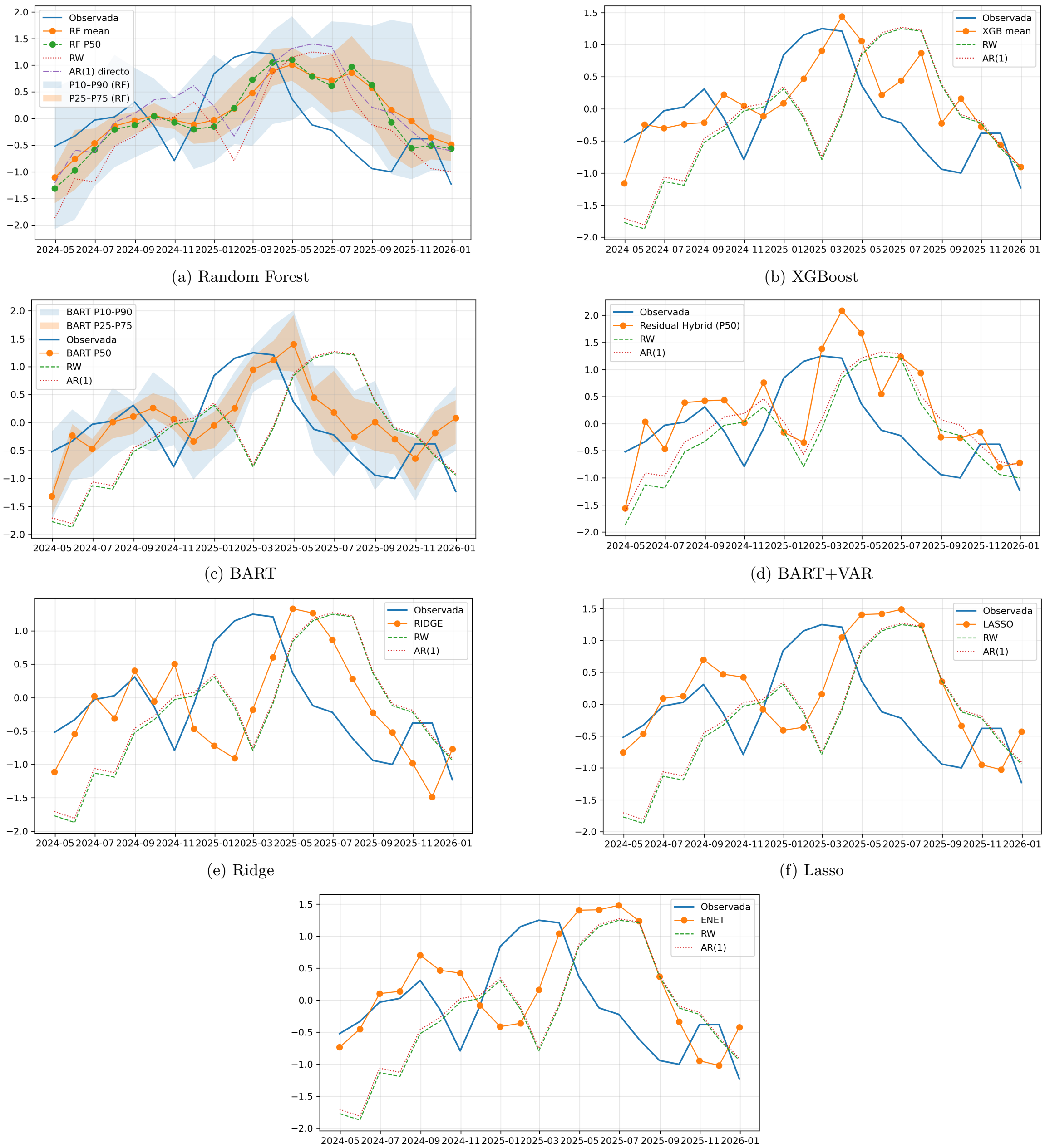
(f) Lasso

(g) Elastic Net

Figure A1: Observed and projected inflation at $t+1$ across models.

Source: Authors' calculations.

Notes: Each panel reports observed year-over-year inflation and model-specific forecasts at horizon $h = 1$.



(a) Random Forest

(b) XGBoost

(c) BART

(d) BART+VAR

(e) Ridge

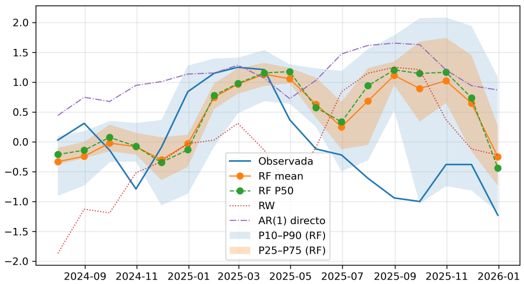
(f) Lasso

(g) Elastic Net

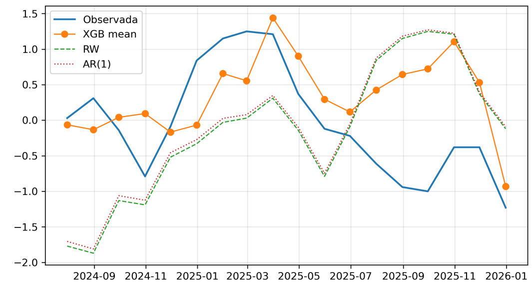
Figure A2: Observed and projected inflation at t+3 across models.

Source: Authors' calculations.

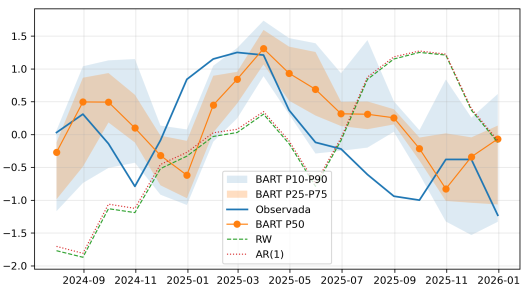
Notes: Each panel reports observed year-over-year inflation and model-specific forecasts at horizon $h = 3$.



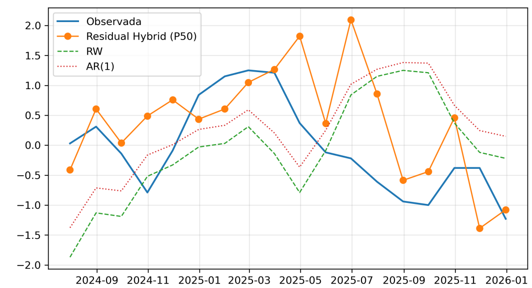
(a) Random Forest



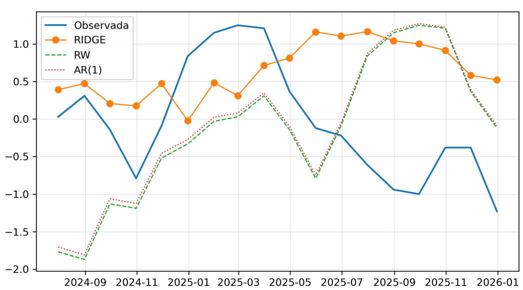
(b) XGBoost



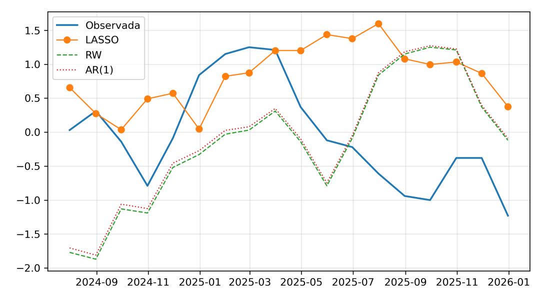
(c) BART



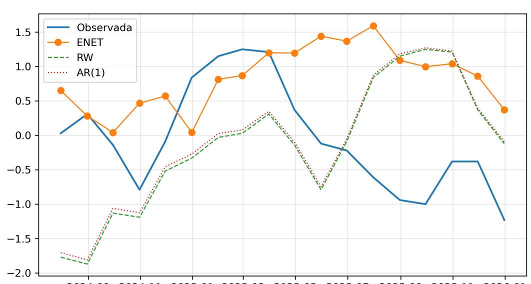
(d) BART+VAR



(e) Ridge



(f) Lasso

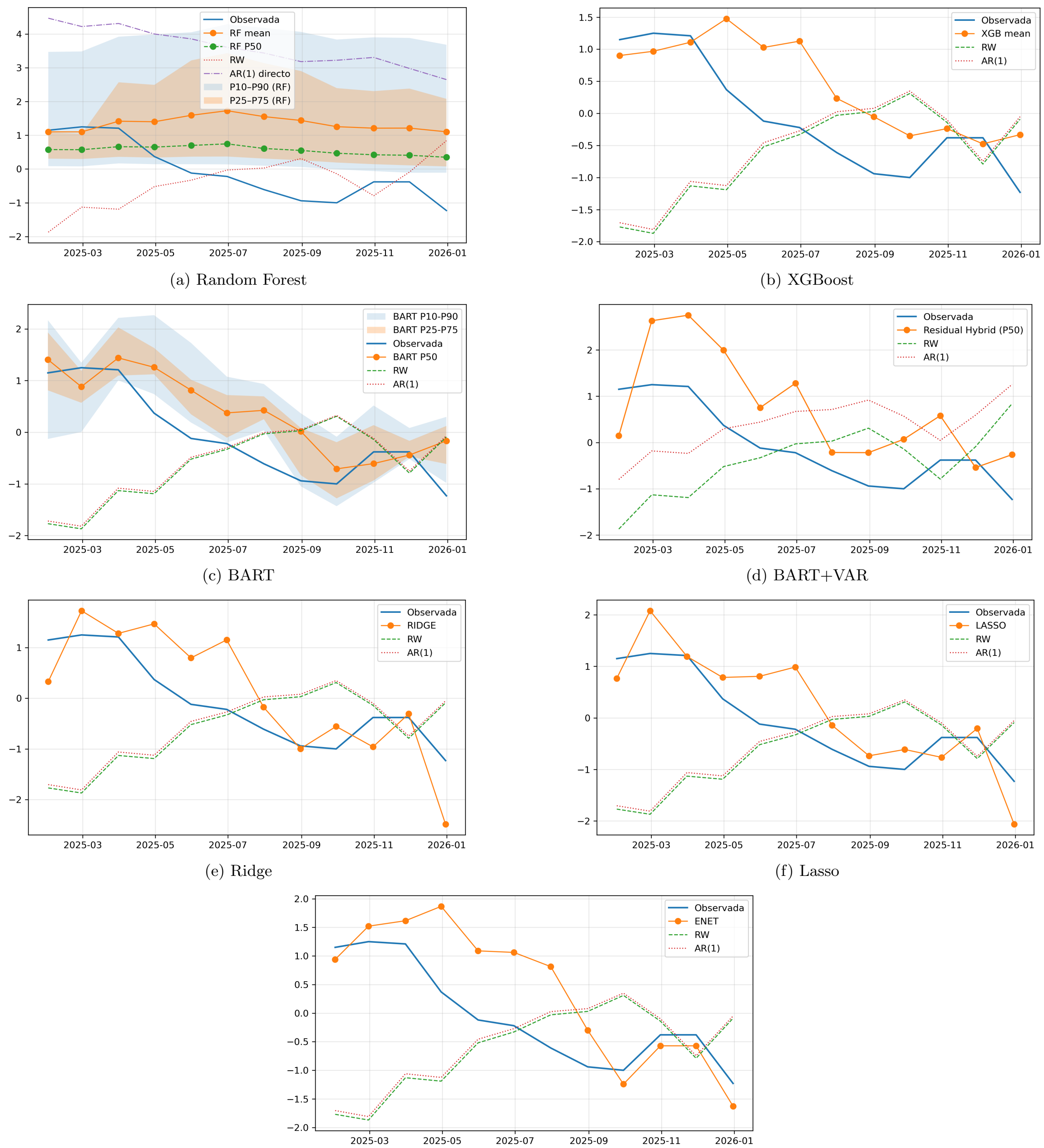


(g) Elastic Net

Figure A3: Observed and projected inflation at $t+6$ across models.

Source: Authors' calculations.

Notes: Each panel reports observed year-over-year inflation and model-specific forecasts at horizon $h = 6$.



(a) Random Forest

(b) XGBoost

(c) BART

(d) BART+VAR

(e) Ridge

(f) Lasso

(g) Elastic Net

Figure A4: Observed and projected inflation at t+12 across models.

Source: Authors' calculations.

Notes: Each panel reports observed year-over-year inflation and model-specific forecasts at horizon $h = 12$.

Table A3: Diebold–Mariano tests (annex): Random Forest vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	RF	RW	SE	2.063	0.039	23.000	2.063	0.039	0.000
t+1	RF	RW	AE	1.862	0.063	23.000	1.862	0.063	0.000
t+1	RF	AR1	SE	2.260	0.024	23.000	2.260	0.024	0.000
t+1	RF	AR1	AE	1.908	0.056	23.000	1.908	0.056	0.000
t+3	RF	RW	SE	−1.287	0.198	21.000	−1.287	0.198	2.000
t+3	RF	RW	AE	−1.630	0.103	21.000	−1.630	0.103	2.000
t+3	RF	AR1	SE	−1.457	0.145	21.000	−1.457	0.145	2.000
t+3	RF	AR1	AE	−1.566	0.117	21.000	−1.566	0.117	2.000
t+6	RF	RW	SE	−2.560	0.011	18.000	−2.084	0.037	2.000
t+6	RF	RW	AE	−1.908	0.056	18.000	−1.671	0.095	2.000
t+6	RF	AR1	SE	−2.468	0.014	18.000	−2.654	0.008	2.000
t+6	RF	AR1	AE	−2.601	0.009	18.000	−2.274	0.023	2.000

Source: Authors' calculations.

Notes: Diebold–Mariano (DM) tests vs benchmarks Random Walk (RW) and AR(1), using loss functions squared error (SE) and absolute error (AE). Baseline Newey–West lag $q = h - 1$; sensitivity uses $q = \min\{h-1, \lfloor (n-1)^{1/3} \rfloor\}$ and reports the resulting lag. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. For $t + 12$ the results are not presented due inconsistencies.

Table A4: Diebold–Mariano tests (annex): XGBoost vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	XGB	RW	SE	−2.736	0.0062	23.000	−2.736	0.0062	0.000
t+1	XGB	AR1	SE	−2.575	0.0100	23.000	−2.575	0.0100	0.000
t+1	XGB	RW	AE	−2.574	0.0100	23.000	−2.574	0.0100	0.000
t+1	XGB	AR1	AE	−2.267	0.0234	23.000	−2.267	0.0234	0.000
t+3	XGB	RW	SE	−3.034	0.0024	21.000	−3.034	0.0024	2.000
t+3	XGB	AR1	SE	−3.078	0.0021	21.000	−3.078	0.0021	2.000
t+3	XGB	RW	AE	−3.049	0.0023	21.000	−3.049	0.0023	2.000
t+3	XGB	AR1	AE	−3.122	0.0018	21.000	−3.122	0.0018	2.000
t+6	XGB	RW	SE	−3.439	0.0006	18.000	−2.874	0.0040	2.000
t+6	XGB	AR1	SE	−3.486	0.0005	18.000	−2.862	0.0042	2.000
t+6	XGB	RW	AE	−3.366	0.0008	18.000	−3.006	0.0026	2.000
t+6	XGB	AR1	AE	−3.456	0.0005	18.000	−2.963	0.0030	2.000
t+12	XGB	RW	SE	−1.554	0.1202	12.000	−1.309	0.1904	2.000
t+12	XGB	AR1	SE	−1.580	0.1142	12.000	−1.316	0.1881	2.000
t+12	XGB	RW	AE	−1.558	0.1193	12.000	−1.186	0.2357	2.000
t+12	XGB	AR1	AE	−1.577	0.1147	12.000	−1.177	0.2394	2.000

Source: Authors' calculations.

Notes: Same conventions as Table A3.

Table A5: Diebold–Mariano tests (annex): BART vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	BART	RW	SE	−2.649	0.0081	23.000	−2.649	0.0081	0.000
t+1	BART	AR1	SE	−2.591	0.0096	23.000	−2.591	0.0096	0.000
t+1	BART	RW	AE	−3.232	0.0012	23.000	−3.232	0.0012	0.000
t+1	BART	AR1	AE	−3.067	0.0022	23.000	−3.067	0.0022	0.000
t+3	BART	RW	SE	−2.537	0.0112	21.000	−2.537	0.0112	2.000
t+3	BART	AR1	SE	−2.497	0.0125	21.000	−2.497	0.0125	2.000
t+3	BART	RW	AE	−2.378	0.0174	21.000	−2.378	0.0174	2.000
t+3	BART	AR1	AE	−2.384	0.0171	21.000	−2.384	0.0171	2.000
t+6	BART	RW	SE	−2.943	0.0033	18.000	−2.410	0.0159	2.000
t+6	BART	AR1	SE	−2.725	0.0064	18.000	−2.297	0.0216	2.000
t+6	BART	RW	AE	−3.545	0.0004	18.000	−2.638	0.0083	2.000
t+6	BART	AR1	AE	−3.244	0.0012	18.000	−2.487	0.0129	2.000
t+12	BART	RW	SE	−1.650	0.0990	12.000	−1.452	0.1465	2.000
t+12	BART	AR1	SE	−1.663	0.0963	12.000	−1.457	0.1451	2.000
t+12	BART	RW	AE	−1.776	0.0757	12.000	−1.441	0.1496	2.000
t+12	BART	AR1	AE	−1.784	0.0745	12.000	−1.432	0.1521	2.000

Source: Authors' calculations.

Notes: Same conventions as Table A3.

Table A6: Diebold–Mariano tests (annex): BART+VAR (residual hybrid) vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	BART+VAR	RW	SE	0.013	0.9895	22.000	0.013	0.9895	0.000
t+1	BART+VAR	RW	AE	0.130	0.8966	22.000	0.130	0.8966	0.000
t+1	BART+VAR	AR1	SE	0.193	0.8473	22.000	0.193	0.8473	0.000
t+1	BART+VAR	AR1	AE	0.152	0.8789	22.000	0.152	0.8789	0.000
t+3	BART+VAR	RW	SE	−1.144	0.2528	21.000	−1.144	0.2528	2.000
t+3	BART+VAR	RW	AE	−1.061	0.2888	21.000	−1.061	0.2888	2.000
t+3	BART+VAR	AR1	SE	−1.408	0.1592	21.000	−1.408	0.1592	2.000
t+3	BART+VAR	AR1	AE	−1.130	0.2585	21.000	−1.130	0.2585	2.000
t+6	BART+VAR	RW	SE	−2.195	0.0282	18.000	−1.374	0.1696	2.000
t+6	BART+VAR	RW	AE	−3.091	0.0020	18.000	−1.600	0.1096	2.000
t+6	BART+VAR	AR1	SE	−1.308	0.1910	18.000	−1.083	0.2787	2.000
t+6	BART+VAR	AR1	AE	−1.978	0.0479	18.000	−1.498	0.1341	2.000
t+12	BART+VAR	RW	SE	−1.584	0.1133	12.000	−1.130	0.2583	2.000
t+12	BART+VAR	RW	AE	−1.033	0.3019	12.000	−0.632	0.5276	2.000
t+12	BART+VAR	AR1	SE	−1.663	0.0963	12.000	−1.195	0.2322	2.000
t+12	BART+VAR	AR1	AE	−1.193	0.2330	12.000	−0.825	0.4096	2.000

Source: Authors' calculations.

Notes: Same conventions as Table A3.

Table A7: Diebold–Mariano tests (annex): Ridge vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	RIDGE	RW	SE	−2.528	0.0115	23.000	−2.528	0.0115	0.000
t+1	RIDGE	AR1	SE	−2.189	0.0286	23.000	−2.189	0.0286	0.000
t+1	RIDGE	RW	AE	−2.084	0.0372	23.000	−2.084	0.0372	0.000
t+1	RIDGE	AR1	AE	−1.714	0.0865	23.000	−1.714	0.0865	0.000
t+3	RIDGE	RW	SE	−0.927	0.3537	21.000	−0.927	0.3537	2.000
t+3	RIDGE	AR1	SE	−0.847	0.3969	21.000	−0.847	0.3969	2.000
t+3	RIDGE	RW	AE	−0.879	0.3796	21.000	−0.879	0.3796	2.000
t+3	RIDGE	AR1	AE	−0.845	0.3982	21.000	−0.845	0.3982	2.000
t+6	RIDGE	RW	SE	−0.838	0.4022	18.000	−0.801	0.4232	2.000
t+6	RIDGE	AR1	SE	−0.810	0.4181	18.000	−0.751	0.4526	2.000
t+6	RIDGE	RW	AE	−0.685	0.4932	18.000	−0.657	0.5112	2.000
t+6	RIDGE	AR1	AE	−0.619	0.5359	18.000	−0.576	0.5649	2.000
t+12	RIDGE	RW	SE	−1.603	0.1091	12.000	−1.385	0.1661	2.000
t+12	RIDGE	AR1	SE	−1.633	0.1026	12.000	−1.396	0.1627	2.000
t+12	RIDGE	RW	AE	−1.808	0.0707	12.000	−1.395	0.1630	2.000
t+12	RIDGE	AR1	AE	−1.843	0.0653	12.000	−1.385	0.1659	2.000

Source: Authors' calculations.

Notes: Same conventions as Table A3.

Table A8: Diebold–Mariano tests (annex): LASSO vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	LASSO	RW	SE	−2.460	0.0139	23.000	−2.460	0.0139	0.000
t+1	LASSO	AR1	SE	−2.369	0.0178	23.000	−2.369	0.0178	0.000
t+1	LASSO	RW	AE	−2.943	0.0033	23.000	−2.943	0.0033	0.000
t+1	LASSO	AR1	AE	−2.723	0.0065	23.000	−2.723	0.0065	0.000
t+3	LASSO	RW	SE	−0.816	0.4145	21.000	−0.816	0.4145	2.000
t+3	LASSO	AR1	SE	−0.769	0.4417	21.000	−0.769	0.4417	2.000
t+3	LASSO	RW	AE	−0.781	0.4346	21.000	−0.781	0.4346	2.000
t+3	LASSO	AR1	AE	−0.760	0.4474	21.000	−0.760	0.4474	2.000
t+6	LASSO	RW	SE	−0.299	0.7649	18.000	−0.282	0.7781	2.000
t+6	LASSO	AR1	SE	−0.237	0.8129	18.000	−0.217	0.8285	2.000
t+6	LASSO	RW	AE	−0.466	0.6414	18.000	−0.424	0.6716	2.000
t+6	LASSO	AR1	AE	−0.399	0.6900	18.000	−0.351	0.7257	2.000
t+12	LASSO	RW	SE	−1.793	0.0730	12.000	−1.557	0.1194	2.000
t+12	LASSO	AR1	SE	−1.835	0.0665	12.000	−1.579	0.1143	2.000
t+12	LASSO	RW	AE	−2.096	0.0361	12.000	−1.663	0.0963	2.000
t+12	LASSO	AR1	AE	−2.146	0.0319	12.000	−1.664	0.0961	2.000

Source: Authors' calculations.

Notes: Same conventions as Table A3.

Table A9: Diebold–Mariano tests (annex): Elastic Net vs benchmarks.

Horizon	Model	Benchmark	Loss	DM stat	p-value	n	DM stat (sens)	p-value (sens)	NW lag (sens)
t+1	Elastic Net	RW	SE	−2.457	0.0140	23.000	−2.457	0.0140	0.000
t+1	Elastic Net	AR1	SE	−2.366	0.0180	23.000	−2.366	0.0180	0.000
t+1	Elastic Net	RW	AE	−2.946	0.0032	23.000	−2.946	0.0032	0.000
t+1	Elastic Net	AR1	AE	−2.724	0.0065	23.000	−2.724	0.0065	0.000
t+3	Elastic Net	RW	SE	−0.821	0.4115	21.000	−0.821	0.4115	2.000
t+3	Elastic Net	AR1	SE	−0.775	0.4383	21.000	−0.775	0.4383	2.000
t+3	Elastic Net	RW	AE	−0.785	0.4327	21.000	−0.785	0.4327	2.000
t+3	Elastic Net	AR1	AE	−0.763	0.4454	21.000	−0.763	0.4454	2.000
t+6	Elastic Net	RW	SE	−0.302	0.7624	18.000	−0.286	0.7745	2.000
t+6	Elastic Net	AR1	SE	−0.241	0.8099	18.000	−0.222	0.8247	2.000
t+6	Elastic Net	RW	AE	−0.467	0.6402	18.000	−0.427	0.6692	2.000
t+6	Elastic Net	AR1	AE	−0.401	0.6884	18.000	−0.354	0.7231	2.000
t+12	Elastic Net	RW	SE	−1.545	0.1224	12.000	−1.242	0.2143	2.000
t+12	Elastic Net	AR1	SE	−1.569	0.1167	12.000	−1.243	0.2138	2.000
t+12	Elastic Net	RW	AE	−1.655	0.0979	12.000	−1.151	0.2497	2.000
t+12	Elastic Net	AR1	AE	−1.673	0.0943	12.000	−1.136	0.2559	2.000

Source: Authors' calculations.

Notes: Same conventions as Table A3.

9.1 Interpretation of XGBoost forecasts: SHAP Analysis

In order to give some interpretation to the results presented for the XGBoost model, we rely on SHAP values, introduced by Lundberg and Lee (2017). SHAP builds on cooperative game theory and attributes to each predictor the marginal contribution it makes to the model prediction relative to a baseline. In general, for a given forecast $\hat{y}_{t+h} = f(Z_t)$, SHAP provides a decomposition of the form

$$f(Z_t) = \phi_0 + \sum_{j=1}^{K} \phi_j,$$

where ϕ_0 is the average prediction over the training sample and ϕ_j is the Shapley value associated with predictor j. Each ϕ_j measures the average change in the prediction when variable j is added to all possible subsets of predictors. In the contexts of models such as XGBoost, SHAP values can be computed to provide a magnitude of the contribution of each covariate.

Table A10: Top 5 predictors by SHAP importance (gain) in the XGBoost model, by forecast horizon

Horizon	Type	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
H+1	Variable	`inf_int_L0`	`ismr_L0`	`ismr_L1`	`ipp_man_L1`	`exp_inf_12_prom_L0`
	Gain	81.6442	55.7572	52.6134	28.0876	13.3873
H+3	Variable	`ipp_man_L1`	`ipp_man_L3`	`ipp_man_L0`	`imp_rd_L3`	`tcn_var_L3`
	Gain	47.9555	10.0379	8.8669	8.2601	6.8065
H+6	Variable	`exp_rd_L6`	`imp_tot_L6`	`tcn_var_L1`	`tbp_L3`	`prime_L1`
	Gain	36.1115	12.6159	10.8213	10.2540	9.7274
H+12	Variable	`tpm_L3`	`exp_rd_L1`	`ip_mpi_L1`	`imp_rd_L0`	`exp_rd_L0`
	Gain	244.0101	103.5469	52.8650	52.5438	49.8373

Notes: Variable names correspond to the original dataset notation.
Notes: Values correspond to SHAP feature importance measured by gain in the final out-of-sample XGBoost model for each forecast horizon.
Source: Authors' calculations.

Table A10 reports the five most influential predictors by gain for the final OOS XGBoost model at each horizon. At the one-month horizon (H+1), current inflation (`inf_int_L0`) dominates, followed by short-run minimum wages (`ismr`) and producer prices (`ipp_man`), suggesting that near-term inflation is strongly driven by persistence and costs of production. Inflation expectations (`exp_inf_12_prom`) also enter among the top predictors, suggesting some forward-looking price-setting behavior.

At H+3 horizon the relevance shifts towards producer prices at different lags, imports of goods and exchange rate variation, which suggests that cost-push and imported inflation channels gain importance as the forecast horizon expands.

Later, at H+6 horizon, the relevance of variables related to trade of goods and exchange rate variation increases but other variables emerges: passive basic rate (deposit rate) and an external rate. This may be reflecting the transmission of financial conditions into inflation dynamics with a lag.

At the annual horizon (H+12), the monetary policy rate (`tpm_L3`) becomes the single most influential predictor, with a markedly higher gain than other variables. This result is economically intuitive: at longer horizons, inflation forecasts are more strongly anchored by monetary policy stance and broader macro-financial fundamentals, rather than short-term persistence.

Overall, the SHAP analysis allows to infer a horizon-dependent structure in inflation dynamics: the shorter-run forecasts are dominated by inertia and cost channels, short-run projections reinforces cost channels and incorporate the external channel, and long-run forecasts are more linked to monetary policy and macroeconomic fundamentals.