



Inflation in Croatia: a new era of forecasting with machine learning

JAKOV ČORAK, univ. mag. oec.*
MIHAEL BRUSAN, student*

Article**

JEL: E37

<https://doi.org/10.3326/pse.50.1.3>

* We express our sincere gratitude to Associate Professor Silvijja Vlah Jerić, PhD, for her exceptional dedication and generosity, as well as for the time, motivation, and invaluable guidance she devoted to helping us complete this research. We are also grateful to Assistant Professor Ivana Lolić, for her contribution, together with Associate Professor Silvijja Vlah Jerić, to the modernization of the study program at the Faculty of Economics and Business in Zagreb through the introduction of the course Applied Machine Learning, which provided the foundational knowledge necessary for the development of this research. We further thank two anonymous reviewers for their constructive comments, which significantly improved the quality of this paper.

The article was judged the best student article in the 2025 annual competition of the Hanžeković Foundation.

** Received: June 1, 2025

Accepted: October 14, 2025

Jakov ČORAK

Croatian National Bank, Trg hrvatskih velikana 3, 10000 Zagreb, Croatia

mail: jakov.corak@gmail.com

ORCID: 0009-0000-8712-9162

Mihael BRUSAN

Faculty of Economics and Business Zagreb, Trg J.F. Kennedy 6, 10000 Zagreb, Croatia

mail: brusanmihael@gmail.com

ORCID: 0009-0007-6543-0764



This is an Open Access article distributed under a Creative Commons Attribution-NonCommercial 4.0 International License which permits non commercial use and redistribution, as long as you give appropriate credit, provide a link to the license, and indicate if changes were made.

Abstract

This paper examines the use of machine learning methods for forecasting inflation in Croatia. Out-of-sample forecasts are generated for multiple horizons using ten models and four alternative sets of input features comprising lags of the target variable, conventional macroeconomic indicators, unconventional variables including Google Trends data, and a combined feature set. Forecast accuracy is assessed across models and relative to a benchmark for the full sample as well as for the periods before and after March 2020 (COVID-19). The results indicate that no single model consistently outperforms others across all settings; however, machine learning methods, particularly tree-based models, deliver superior performance under specific conditions. The forecasts produced by the two best-performing models, SARIMA and LightGBM, exceed the accuracy of the European Commission's projections. As the first paper to apply machine learning to inflation forecasting in Croatia, this study introduces modern analytical techniques into the Croatian forecasting literature.

Keywords: inflation, machine learning, forecasting, macroeconomics, Croatia

1 INTRODUCTION

After decades of low and stable inflation in developed economies, galloping inflation seemed almost unimaginable. In extreme cases, some economists considered it a phenomenon of the past. However, due to recent shocks, inflation has once again emerged as one of the key macroeconomic issues, spurring intense interest from both the academic and professional communities. In this context, successful inflation forecasting becomes crucial for economic policymakers, especially for central banks implementing monetary policy. Timely and reliable inflation forecasts are essential for making adequate monetary policy decisions, such as those related to changing key interest rates, primarily because of the medium-term nature of the monetary transmission mechanism. For this reason, central banks must act in an anticipatory manner and rely primarily on projections of future inflation developments.

Inflation forecasting is a challenging task, as confirmed by the large forecast errors produced by models used by many central banks and international institutions (Medeiros et al., 2021: 100). In recent years, machine learning (ML) models have gained importance due to their numerous advantages over traditional econometric models used in macroeconomic modelling. The big data environment naturally complements ML methods because it allows for the exploitation of their ability to process a large number of features and uncover complex relationships within the data.

The main goal of this paper is to determine which ML model is the most successful in out-of-sample inflation forecasting in Croatia. Particular emphasis is placed on comparing the predictive capabilities of ML models against traditional econometric time series approaches, such as SARIMA. The paper further examines the forecasting performance of the models under different conditions, before and after the COVID crisis. The pandemic, a strong exogenous shock, caused high levels of

volatility and uncertainty, producing an environment that the models, trained during more stable macroeconomic periods (with the exception of the 2008 crisis), had not previously encountered. Separating less volatile from more volatile periods, therefore, shows whether the models can maintain accuracy in a changed macroeconomic environment and how informative conventional versus unconventional features are in stable as compared to stressful times. Finally, the best model is selected and its forecasts are compared with the forecasts of the European Commission.

This paper contributes to the literature in at least two ways. First, the paper investigates the application of modern ML methods to inflation forecasting in Croatia and constitutes, to the best of the authors' knowledge, the first contribution of its kind in the Croatian literature. Second, the comparison of inflation forecasts across different economic regimes, using various models and features, provides a wealth of information regarding the predictive capabilities of the models, as well as the informativeness and importance of the features employed.

The remainder of the paper is organized as follows. The second chapter presents an extensive review of the international and domestic literature on inflation forecasting using various models, with a particular focus on ML models. The third chapter includes a description of all variables used, including an explanation of their categorization, and sets out the methodological framework, focusing on model selection, data preparation and transformation, the forecasting procedure, and measures of forecast accuracy (the prediction error indicators used). Chapter four presents the results, namely the forecast errors of all utilized models, and selects the best models, whose predictions are then compared with those of the European Commission. Fifth chapter discusses the results. The conclusion is provided at the end of the paper.

2 LITERATURE REVIEW

The modelling of inflation is one of the fundamental issues in macroeconomic analysis and has been the subject of various theoretical and empirical approaches over decades. The typical framework for understanding inflation dynamics long relied on the concept of the Phillips curve. Although Blinder (1997: 241) emphasizes that the Phillips curve served as a reliable tool in inflation forecasting for decades, later works question its predictive power. For instance, Atkeson and Ohanian (2001) demonstrate that the Phillips curve in the US does not even outperform simple naive models in forecasting inflation and conclude that “the search for yet another Phillips curve-based inflation forecasting model should be abandoned”. Similar findings are confirmed by Ang, Bekaert and Wei (2005), who show that ARMA models and, in particular, forecasts based on surveys, achieve lower forecast errors than Phillips curve models.

The poor empirical forecasting performance of traditional theoretical models has prompted researchers to experiment with different sets and types of variables, as well as various model approaches, in order to improve the accuracy of inflation forecasts. In this context, the literature shows a growing interest in including

additional information in models to enhance the accuracy of inflation forecasts. For example, Chen, Turnovsky and Zivot (2014) include aggregates of world commodity prices in models, Forni et al. (2003), Stock and Watson (2003), and Monteforte and Moretti (2013) use financial variables, while Groen, Paap and Ravazzolo (2013) and Ang, Bekaert and Wei (2005) use expectation variables. As the set of relevant predictors expanded, researchers often resorted to factor models, which compress information from a multitude of variables into a few latent components – factors (e.g. Eickmeier and Ziegler, 2008). However, their efficiency largely depends on the quality of the variables used in constructing the factors. Some research indicates that expanding the database does not necessarily result in better forecasting outcomes (e.g. Barhoumi, Darné and Ferrara, 2009).

In the last decade, ML models have received special attention in the literature. Unlike factor models, which reduce dimensionality by summarizing information into a few components, ML models allow for the direct inclusion of a large number of features. Although the application of ML to inflation forecasting is still relatively new, numerous empirical studies in the international literature already confirm its advantages over traditional econometric models. Nakamura (2004) shows that neural networks outperform a simple univariate AR model in forecasting US inflation at short horizons. Subsequent studies confirm the superior performance of ML and regularized models to that of traditional linear benchmarks. In particular, Medeiros and Mendes (2016) and Garcia, Medeiros and Vasconcelos (2017) find that LASSO-type models dominate standard AR and factor models in forecasting inflation in the US and Brazil, respectively. Using a broader set of ML techniques, Medeiros et al. (2021) report that Random forest (RF) delivers the lowest forecast errors for US inflation and performs robustly across different economic conditions, a result also supported by Ülke, Sahin and Subasi (2018) for more volatile time series. However, evidence suggests that this superiority is not uniform across time. Naghi, O’Neill and Zaharieva (2024) show that while RF outperforms benchmark models prior to COVID-19, its performance deteriorates during the pandemic and high-inflation period, when SVM and GBM yield more accurate forecasts. Overall, the literature indicates that ML methods tend to outperform ARMA-type benchmarks (Araujo and Gaglianone, 2023), particularly at longer horizons where tree-based models such as RF and XGBoost perform especially well, although simpler autoregressive models may remain competitive in smaller samples and data-constrained economies (Ivaşcu, 2023).

Inflation has received limited attention in the Croatian economic literature, as most studies have concentrated on explaining inflation dynamics and their links with other macroeconomic variables (e.g. Payne, 2002), with little emphasis placed on inflation forecasting. This gap in the literature constitutes the main motivation for this paper. To the authors’ best knowledge, only two studies explicitly examine inflation forecasting in Croatia. Pufnik and Kunovac (2006) use a SARIMA model to forecast short-term inflation and show that, over longer horizons, aggregating forecasts of individual CPI sub-components improves accuracy

relative to directly forecasting the aggregate CPI. Kunovac (2007) applies principal component analysis and finds that information extracted from a large set of macroeconomic variables enhances forecasting performance, with even a single factor outperforming benchmark models.

3 METHODOLOGY AND DATA

Although numerous definitions of machine learning exist in the literature, this paper relies on the definition provided by Masini, Medeiros and Mendes (2021: 77), according to which: “machine learning is the combination of automated computer algorithms with powerful statistical methods to learn (discover) hidden patterns in rich datasets”. This definition highlights the key characteristic of ML: the model’s ability to autonomously learn complex patterns from a large amount of data, including nonlinearities and interactions among variables.

To forecast inflation in Croatia, a total of ten linear and non-linear models were employed, utilizing the traditional SARIMA model as a primary benchmark. The forecasting success of each model is also compared with a naive model’s forecast using the slightly adjusted MASE forecast error, which is introduced later.

3.1 DATA

The target variable is the month-on-month inflation rate, measured as the monthly change in the Harmonised Index of Consumer Prices (HICP). The model specification includes lagged inflation and ten additional variables which, after applying the transformations described in section 3.2, give rise to several hundred input features. This approach aims to maximise forecasting accuracy while remaining feasible given the available computational resources.

TABLE 1

Input features

Variable	Feature label	Source
Lagged target variable	<i>laghicip</i>	Eurostat
Unemployment rate	<i>unemp</i>	CBS
Industrial production volume index	<i>ind</i>	CBS
Real retail trade turnover, annual change rates (seasonally and calendar-adjusted)	<i>retail</i>	CBS
US 10-year and 2-year treasury yield spread	<i>spread</i>	FRED
Average Brent and WTI crude oil prices	<i>bwti</i>	EIA
Baltic exchange dry index	<i>badi</i>	Bloomberg
Consumer confidence index	<i>conf</i>	CNB
Economic policy uncertainty index	<i>epu</i>	EPU (Sorić and Lolić 2017)
Search popularity of “Urlaub in Kroatien” (Vacation in Croatia) in Germany	<i>gturlaub</i>	Google Trends
Search popularity of the term “Inflation” in Germany	<i>gtinfl</i>	Google Trends

Source: Authors’ own elaboration.

Table 1 provides an overview of the input features. The first variable (*laghicp*) captures lagged inflation. The remaining variables are grouped into *conventional* and *unconventional* features. The former correspond to the variables listed in rows two to seven and represent macroeconomic indicators commonly used in inflation modelling (hard data). The latter correspond to the last four rows and include high-frequency indicators and sentiment measures (soft data) that provide timely information on current economic conditions and expectations and are widely used in contemporary modelling.

3.2 FEATURE TRANSFORMATION

The time series used in the analysis begin on 31 January 2006 and were retrieved on 31 March 2025. To replicate authentic forecasting conditions and prevent data leakage, missing values caused by publication delays (e.g., unemployment data) were filled by shifting the corresponding variables forward by one month. Following this alignment, an extensive feature engineering process was applied to the 11 base variables using five distinct transformation methods:

- 1) Rolling mean: the average of the previous six months is computed, which reduces short-term volatility and emphasises medium-term trends.
- 2) Exponential moving average (EMA): unlike the rolling mean, the EMA assigns greater weight to more recent observations, meaning that newer features are of greater importance in the model than older ones.
- 3) First differences: obtained by subtracting consecutive observations ($y_t - y_{t-1}$), which removes the trend component and transforms the series into stationary processes.
- 4) Lags: for each variable, lags from one to nine months were generated, as some of the utilized variables have a proven forecasting property several months ahead.
- 5) Standardization: every feature was converted to a form with a mean of zero and a standard deviation of one.

These transformations expanded the initial input space to a maximum of 440 features. Consequently, due to the initialization period required for lags and moving averages, the final usable dataset consists of 215 observations.

3.3 MODELS

The forecast is given by the equation:

$$\pi_{t+h} = G_h(x_t) + u_{t+h}, \quad h = 1, \dots, H; \quad t = 1, \dots, T \quad (1)$$

where π_{t+h} is the inflation in month $t + h$, $x_t = (x_{t1}, \dots, x_{tn})^T$ is the vector of input data (features). $G_h(\cdot)$ is the function that connects the input data with inflation, and u_{t+h} is a random error with a mean value of zero (Medeiros et al., 2021). The goal of the model is to estimate the function G_h that minimizes the forecast error.

Ridge regression (RR) was proposed by Hoerl and Kennard (1970). In essence, it is a linear regression model that differs from the usual OLS only in the estimation of the coefficient parameter. The sum of the least squares of deviations is

minimized, as is standard, with an added penalty term applied to the error function. The penalty function is defined as:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda) := \lambda \sum_{i=1}^n \beta_{h,i}^2 \quad (2)$$

where $\beta_{h,i}$ is the i -th parameter for horizon h , λ is the regularization factor which, as a hyperparameter, determines the strength of the penalization, and the penalty function $p(\beta_{h,i}; \lambda)$ assigns λ for the square of each coefficient β (Medeiros et al., 2021).

Lasso was proposed by Tibshirani (1996). Similar to the RR model, it is a linear regression model. Unlike the former (L2), it uses L1 penalization, which can be defined as follows:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda) := \lambda \sum_{i=1}^n |\beta_{h,i}| \quad (3)$$

where the notation is read similarly to that for ridge regression. Lasso performs simultaneous regularization and variable selection. Less informative features receive a coefficient equal to zero, meaning the model discards features that do not contribute to inflation forecasting.

The Elastic net (ENet) model, developed by Zou and Hastie (2005), combines the RR and Lasso approaches. In other words, the ENet model uses L1 and L2 penalization for balanced regularization in the following way:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda, \alpha) := \alpha \lambda \sum_{i=1}^n \beta_{h,i}^2 + (1 - \alpha) \lambda \sum_{i=1}^n |\beta_{h,i}| \quad (4)$$

where $\alpha \in [0, 1]$ (Medeiros et al., 2021).

Support vector regression (SVR), introduced in the paper by Drucker et al. (1997) is a regularised regression approach that can capture nonlinear relationships through a feature mapping $\phi(\cdot)$ (or, equivalently, a kernel). In SVR, the model is chosen to be as flat as possible while allowing errors within an ε -insensitive tube, which leads to the optimisation problem:

$$\min_{w, b, \zeta, \zeta^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\zeta_i + \zeta_i^*) \text{ s.t. } \begin{cases} y_i - (w^\top \phi(x_i) + b) \leq \varepsilon + \zeta_i \\ (w^\top \phi(x_i) + b) - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \quad (5)$$

Here, ε sets the width of the tube, while C controls how strongly violations outside the tube are penalised via the slack variables ζ_i, ζ_i^* . In practice, the problem is solved by moving to the dual formulation and solving a quadratic programming problem, as described in the paper. The resulting predictor depends only on a subset of training observations (the “support vectors”), which makes the method particularly attractive in high-dimensional settings. In forecasting applications, SVR is then applied by tuning ε , C , and kernel parameters (e.g., via cross-validation) and evaluating the fitted function on new inputs to generate predictions.

The Random Forest (RF) is an ensemble of multiple simple decision trees designed to reduce the variance of individual regression trees. It is created by combining a large number of randomly constructed trees (Breiman, 2001). A regression tree is in itself a nonparametric model that approximates an unknown nonlinear function using local predictions and the recursive partitioning of the covariate space (Breiman, 1996). The RF for a regression problem is defined as an ensemble of M randomly generated trees. For the j -th tree in the ensemble, the predicted value at query point x is denoted by $m_n(x; \Theta_j, D_n)$, where Θ_j represents independent random variables controlling the construction process of individual trees, and $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ represents the training sample. The goal is to use the dataset D_n to construct an estimate $m_n : [0, 1]^p \rightarrow \mathbb{R}$ of the function m . The estimate of the RF with a finite number of trees is then given by the expression:

$$m_{M,n}(x; \Theta_1, \dots, \Theta_M, D_n) = \frac{1}{M} \sum_{j=1}^M m_n(x; \Theta_j, D_n) \quad (6)$$

(Scornet, Biau and Vert, 2015). Since the number of trees can be made very large in practice, it's natural to consider the case $M \rightarrow \infty$. Then, the RF estimate converges to the expectation of the individual tree predictions (Scornet, Biau and Vert, 2015):

$$m_n(x; D_n) = E_{\Theta} [m_n(x; \Theta, D_n)] \quad (7)$$

The Gradient Boosting Model (GBM) constructs a complex predictor by summing shallow regression trees, where each new tree f_m is trained to minimize the residual errors of the previous ensemble (Friedman, 2001). The final model is represented as:

$$F_M(x_t) = F_0(x_t) + \sum_{m=1}^M v f_m(x_t) \quad (8)$$

where v acts as a shrinkage parameter to control overfitting. While the standard GBM effectively captures non-linearities, advanced implementations have been developed to improve efficiency. XGBoost (Extreme Gradient Boosting) optimizes the standard GBM framework by adding explicit regularization terms $\Omega(f)$ and utilizing parallel processing, resulting in more robust generalization (Friedman, 2001; Chen and Guestrin, 2016). For further scalability in high-dimensional environments, LightGBM introduces Gradient-based One-side sampling (GOSS) and Exclusive feature bundling (EFB). These techniques drastically reduce data scanning requirements by focusing on instances with large gradients and bundling sparse features, achieving up to 20-fold faster training speeds with comparable accuracy (Ke et al., 2017).

Neural networks (NNs) are multi-layer systems of interconnected neurons inspired by biological processes, whose purpose is to map input features into forecasts through a sequence of nonlinear transformations. Although NNs are highly effective at modelling complex nonlinear relationships, they are often described as “black-box” models due to the limited interpretability of their parameters. Consider a

multilayer perceptron (MLP) with a total of L layers, where the first $L - 1$ layers are hidden and the L -th layer is the output layer. Let x_t denote the input vector at time t . We define the layer outputs (activations) recursively using the notation z as:

$$\begin{aligned} z_0 &= x_t, \\ z_\ell &= \sigma_\ell (W_\ell z_{\ell-1} + b_\ell), \ell = 1, \dots, L-1 \end{aligned} \quad (9)$$

where W_ℓ and b_ℓ denote the weight matrix and bias vector of layer ℓ , and $\sigma_\ell(\cdot)$ is the nonlinear activation function of that layer (e.g., ReLU). The h -step-ahead forecast is then obtained via an affine projection from the last hidden layer:

$$\hat{\pi}_{t+h} = g(W_L z_{L-1} + b_L) \quad (10)$$

where $g(\cdot)$ is the output-layer function. The set of parameters $\{W_\ell, b_\ell\}_{\ell=1}^L$ is estimated by minimizing a loss function $\mathcal{L}(\pi_{t+h}, \hat{\pi}_{t+h})$. The gradients of the loss with respect to the parameters are computed using the backpropagation algorithm, and the parameters are updated using a chosen optimization procedure (Rumelhart, Hinton and Williams, 1986).

Although traditionally considered a statistical time series model, the Seasonal Autoregressive Integrated Moving Average (SARIMA) is increasingly integrated into modern ML frameworks, often serving as a robust benchmark for automated predictive systems. SARIMA extends the standard ARIMA architecture by explicitly incorporating seasonal components, enabling the effective recognition of periodic patterns. The model is formally defined by the expression:

$$\Phi_p(B^s)\phi_p(B)(1-B^s)^D(1-B)^d\pi_t = \Theta_q(B^s)\theta_q(B)\varepsilon_t \quad (11)$$

where B denotes the lag operator, $(1 - B)$ and $(1 - B^s)$ represent the non-seasonal and seasonal differencing required for stationarity, while the polynomials ϕ and θ (along with their seasonal counterparts) capture the autoregressive and moving average dynamics. Parameter selection is performed by automated algorithms (auto-SARIMA) that optimize the model structure based on information criteria such as the Akaike information criterion (AIC), ensuring both transparency and diagnostic robustness (Hyndman and Athanasopoulos, 2021).

Developed by Facebook's research division, the Prophet model specializes in forecasting time series characterized by pronounced seasonal patterns and occasional anomalies. Unlike traditional autoregressive models, Prophet relies on an additive decomposition methodology that treats the forecasting problem as a curve-fitting exercise, separating the signal into three distinct components:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (12)$$

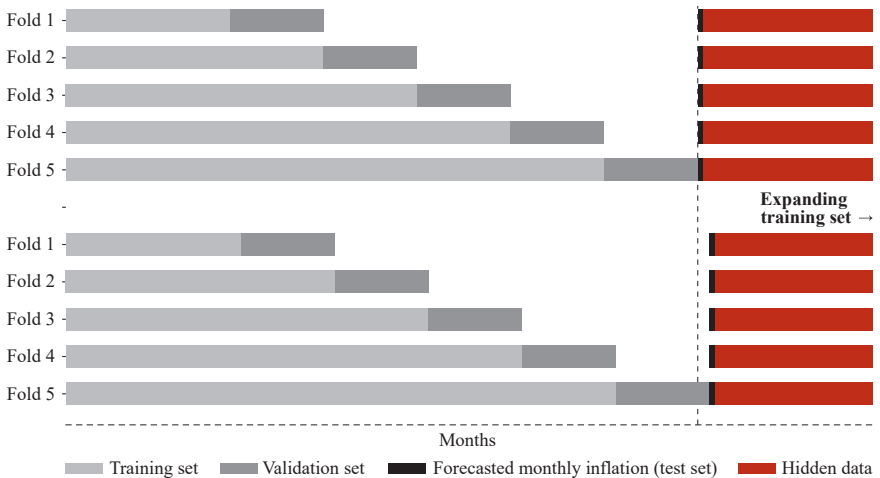
where $g(t)$ represents trend component, $s(t)$ is the seasonal component, $h(t)$ represents the influence of holidays or specific events (Taylor and Letham, 2018; Hyndman and Athanasopoulos, 2021).

3.4 FORECASTING PROCEDURE

To ensure generalization and prevent overfitting, this study employs an expanding window cross-validation technique, specifically adapted for time series data. As illustrated in figure 1, the model is initially trained on a fixed set, which progressively expands by the addition of one new month of data in each iteration. Following each expansion, the model is fully retrained and validated on the subsequent month, a process that meticulously simulates real-world forecasting and eliminates data leakage.

FIGURE 1

Illustrative display of the expanding window approach



Source: Authors' own elaboration.

Model selection within each cross-validation fold is determined by minimizing the Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{t=1}^n (\pi_t - \hat{\pi}_t)^2 \tag{13}$$

Final predictive performance is evaluated using two metrics. The first is the RMSE defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\pi_t - \hat{\pi}_t)^2} \tag{14}$$

The second metric used is the out-of-sample MASE (OMASE), which is nearly identical to the measure proposed by Hyndman and Koehler (2006), with the exception that the denominator here is the out-of-sample MAE of the naive model,

whereas Hyndman and Koehler employ the in-sample MAE of the naive model. OMASE can be written as:

$$OMASE = \frac{\frac{1}{n} \sum_{t=1}^n |\pi_t - \widehat{\pi}_t|}{\frac{1}{n-m} \sum_{t=m+1}^n |\pi_t - \pi_{t-m}|}, m = 12 \quad (15)$$

where m is the length of the seasonal cycle (for monthly data 12). An OMASE value below 1 indicates that the model under evaluation outperforms the naive seasonal benchmark.

To rigorously evaluate forecasting performance and feature importance under varying economic conditions, 12 experimental configurations were developed by crossing three evaluation periods with four distinct input feature sets:

- Evaluation periods:
 - The entire available period.
 - The period up to February 2020.
 - The period from March 2020 onward (the start of the lockdown).
- Input feature combinations:
 - Target variable lags only.
 - Lags + *conventional* features.
 - Lags + *unconventional* features.
 - All available variables combined.

To benchmark performance against that of the European Commission (EC), generated monthly forecasts are aggregated into annual inflation rates (see appendix). The study strictly replicates the data availability constraints of the EC's three key reporting cycles to ensure a fair comparison:

- 1) Spring Forecast (current year): Published in May with an April cut-off. Only information available by the end of April is used, with delayed-release variables imputed using lagged values. Given the availability of a HICP flash estimate for April, eight monthly inflation rates must be forecast, corresponding to horizons $h = 1, \dots, 8$.
- 2) Autumn Forecast (current year): Published in November with an October cut-off. Data availability is aligned analogously, including a flash estimate for October. As ten monthly observations are known, forecasts are generated for the remaining two months ($h = 1, 2$).
- 3) Autumn Forecast (following year): Based on the same October cut-off and information set as the autumn forecast for the current year. Constructing the annual inflation rate for the following year requires forecasting the final two months of the current year and all twelve months of the next year, resulting in fourteen monthly projections ($h = 1, \dots, 14$).

4 RESULTS

This chapter reports the results of all models used to forecast monthly inflation in Croatia. The forecasting setup and evaluation procedure follow the framework outlined in chapter 3, with specific reference to section 3.4. Owing to computational constraints, forecasts are first generated for horizons of 3, 6, and 9 months, after which the best-performing model is employed to produce forecasts for the full set of 14 horizons. Table 2 presents RMSE forecast errors for the period up to March 2020, corresponding to the pre-COVID-19 sample.

TABLE 2

RMSE in the pre-COVID-19 period for the 3-, 6- and 9-month forecasting horizons

Model	Horizon	lag	unconv	conv	all
SARIMA	3	0.2941	/	/	/
RIDGE		0.3233	0.3188	0.3270	0.3133
LASSO		0.3349	0.3470	0.3353	0.3286
ENET		0.3361	0.3448	0.3351	0.3291
PROPHET		0.3509	0.6340	1.6688	0.8964
NN		0.3239	0.3856	0.3267	0.3435
SVR		0.3420	0.3535	0.3580	0.3622
RF		0.3609	0.3741	0.3872	0.3622
XGBOOST		0.3420	0.3166	0.3679	0.3037
LGBM		0.3180	0.3468	0.3286	0.3211
SARIMA	6	0.2874	/	/	/
RIDGE		0.3236	0.3144	0.3253	0.3123
LASSO		0.3404	0.3726	0.3594	0.3637
ENET		0.3499	0.3729	0.3638	0.3596
PROPHET		0.3778	0.6283	2.3085	1.0580
NN		0.3239	0.3543	0.3992	0.3624
SVR		0.3382	0.3530	0.3487	0.4065
RF		0.3409	0.3787	0.4631	0.4065
XGBOOST		0.2993	0.3112	0.3211	0.3270
LGBM		0.2728	0.3428	0.3867	0.3130
SARIMA	9	0.2905	/	/	/
RIDGE		0.3167	0.3096	0.3144	0.3007
LASSO		0.3314	0.3359	0.3356	0.3369
ENET		0.3349	0.3430	0.3365	0.3372
PROPHET		0.3554	0.7249	1.5218	0.7932
NN		0.3071	0.3248	0.3225	0.3337
SVR		0.3405	0.3226	0.3237	0.3585
RF		0.3067	0.3199	0.4429	0.3585
XGBOOST		0.2641	0.2983	0.2892	0.2946
LGBM		0.2962	0.3068	0.2632	0.2804

Note: Darker shading corresponds to lower RMSE values. Bolded entries denote combinations of models and feature sets that achieved OMASE > 1. Non-bolded entries achieved OMASE < 1. In appendix in tables A1-A3 results are presented as indices relative to the SARIMA model.

Source: Authors' calculations.

The results presented in table 2 indicate that tree-based models (LightGBM and XGBoost) achieve above-average accuracy in a period of subdued volatility (the period preceding the outbreak of the COVID-19 pandemic). At longer forecasting horizons (6 and 9 months), they consistently outperform the traditional SARIMA model, with their performance in some cases further enhanced by the inclusion of external features. Among the linear models, only the Ridge model maintains OMASE values below one across all horizons. Given that ENet and Lasso perform substantially worse than Ridge, it can be concluded that the feature exclusion induced by the L1 penalty does not, in this case, contribute to improved forecast accuracy. Prophet is the only model in the pre-COVID-19 period that fails to outperform the simple (naïve) forecasting model under any combination of input features. Moreover, for no other model does the inclusion of external features deteriorate forecasting performance to the same extent as it does for Prophet. Table 2 does not reveal any clear or systematic pattern whereby unconventional variables consistently improve forecasting accuracy relative to conventional variables, or vice versa. Table 3 reports the forecasting errors for the period following the onset of the COVID crisis.

TABLE 3

RMSE in the post-COVID-19 period for the 3-, 6- and 9-month forecasting horizons

Model	Horizon	lag	unconv	conv	all
SARIMA	3	0.5653	/	/	/
RIDGE		0.6081	0.6710	0.6581	0.6362
LASSO		0.5952	0.6332	0.6063	0.6208
ENET		0.5869	0.6342	0.6207	0.6272
PROPHET		0.5774	0.8437	0.9092	3.2657
NN		0.6186	0.7196	0.6576	0.6210
SVR		0.6364	0.6434	0.6059	0.6697
RF		0.6335	0.6281	0.6768	0.6697
XGBOOST		0.6637	0.6773	0.5894	0.5786
LGBM		0.6294	0.5923	0.5716	0.5746
SARIMA		6	0.5836	/	/
RIDGE	0.5862		0.8057	0.6623	0.7214
LASSO	0.6002		0.6542	0.6123	0.6112
ENET	0.5895		0.6507	0.6205	0.6354
PROPHET	0.6112		0.9299	1.2737	3.5227
NN	0.5979		0.9232	0.6839	0.7940
SVR	0.6343		0.6690	0.6401	0.6962
RF	0.6559		0.6527	0.7164	0.6962
XGBOOST	0.6622		0.7050	0.6103	0.6797
LGBM	0.6603		0.6029	0.6321	0.6086
SARIMA	9		0.5792	/	/
RIDGE		0.5745	0.8741	0.6380	0.7602
LASSO		0.5882	0.5936	0.6274	0.6055
ENET		0.5862	0.6032	0.6272	0.6087
PROPHET		0.5946	1.1732	1.0935	7.5208
NN		0.5963	1.0482	0.6661	0.8043
SVR		0.6323	0.6498	0.6482	0.6812
RF		0.6223	0.6556	0.6945	0.6812
XGBOOST		0.6438	0.6536	0.6372	0.6368
LGBM		0.6349	0.5886	0.5953	0.6082

Source: Authors' calculations.

Table 3 shows that the SARIMA model consistently achieves the lowest RMSE across all forecasting horizons. Among the ML models, the closest competitor to SARIMA is LightGBM. At shorter horizons, LGBM particularly benefits from the inclusion of conventional variables, whereas at longer horizons it attains higher predictive accuracy when unconventional features are included. Among the linear models, the Ridge model without external variables proves to be the most stable across horizons, although Lasso and ENet fall within a comparable accuracy range. Prophet records OMASE values below 1 only when it relies exclusively on lags of the target variable, which suggests that the expanded feature set offers insufficient informational value for this model. Neural networks also perform best when only temporal lags of the target variable are used, indicating that they have limited ability to extract useful information from additional external predictors. Furthermore, tables 2 and 3 reveal that OMASE values fall below 1 more frequently in the post-COVID than in the pre-pandemic period. This implies that, after the onset of the pandemic, models more often outperformed the naïve benchmark in terms of forecast accuracy. At the same time, table 3 shows that RMSE values for all models are considerably higher than those reported in table 2, pointing to increased unpredictability of inflation in the post-pandemic environment. Heightened volatility, structural changes in the economy, and global shocks such as supply chain disruptions and the sharp rise in energy prices all contributed to this increased uncertainty, ultimately reducing the predictive accuracy of all models under consideration. Finally, table 4 reports forecasting errors for the full sample period.

The results in table 4 for the full observation period reveal several important patterns regarding model performance. First, SARIMA attains the lowest forecasting errors across the selected horizons. Second, among the linear models, the lowest forecast errors are most frequently obtained when only temporal lags of the target variable are used as input features. Third, among the nonlinear models, LightGBM with external variables exhibits consistent predictive superiority over the other models. Across multiple feature combinations and all horizons, it records low RMSE and OMASE values, often the lowest among all models except SARIMA. Alongside LightGBM, XGBoost also stands out among nonlinear models, although its predictive performance is somewhat less consistent across different sets of external variables.

To enable a comparison between the best-performing model and the European Commission's forecasts in the final stage of the analysis, we proceed to select a model that will serve as the representative model for the final forecast and comparison. Since it is difficult to establish a single criterion to unambiguously identify the most successful model, the most appropriate approach appears to be focusing on the consistency of each model's results across different macroeconomic regimes.

Given that SARIMA achieves the lowest forecasting errors in most cases, across all horizons and all periods examined, it can be regarded as the best-performing model for forecasting inflation in Croatia. If SARIMA is excluded as a traditional time-series benchmark, LightGBM emerges as the most consistent performer among the ML models. LightGBM regularly attains low forecasting errors in all periods considered and across nearly all combinations of input variables (see appendix table A4).

TABLE 4

RMSE over the full observation period for the 3-, 6- and 9-month forecasting horizons

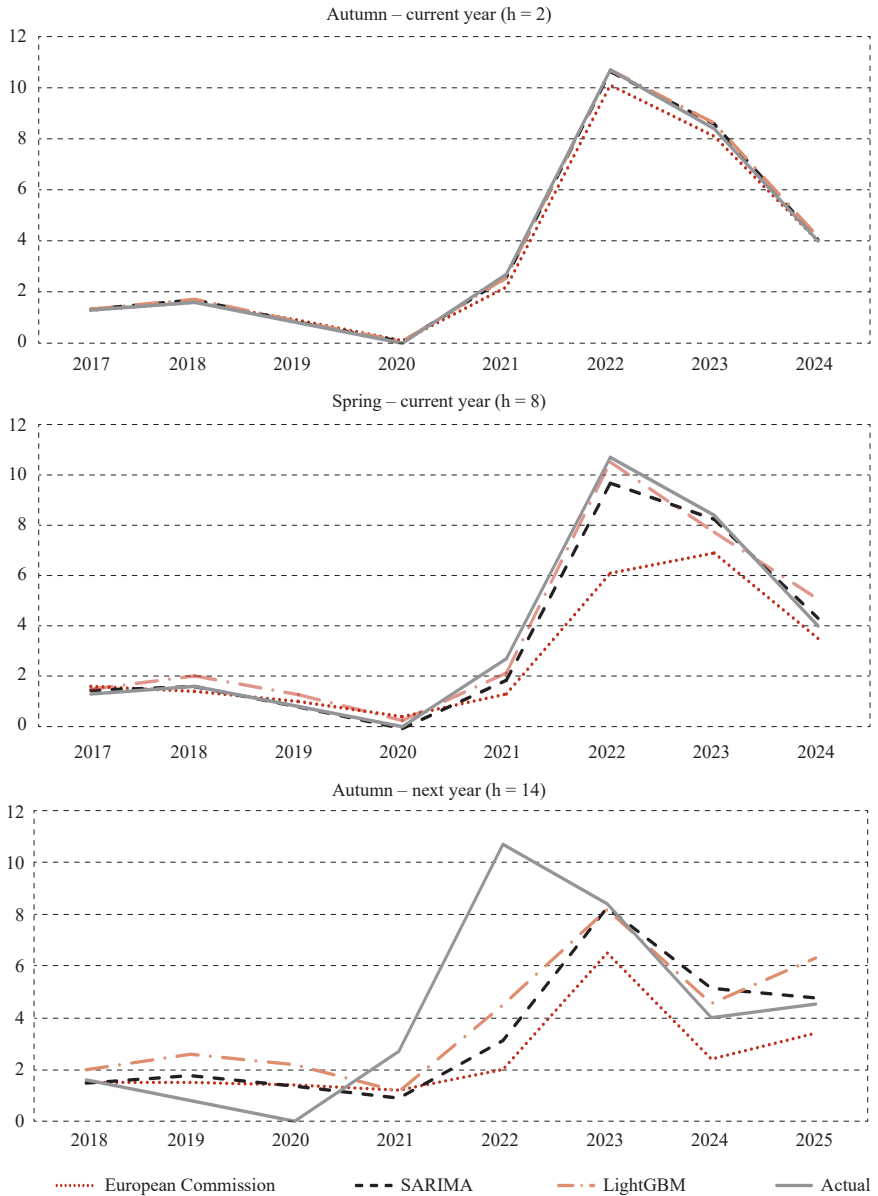
Model	Horizon	lag	unconv	conv	all
SARIMA	3	0.4772	/	/	/
RIDGE		0.5161	0.5606	0.5531	0.5341
LASSO		0.5124	0.5390	0.5194	0.5283
ENET		0.5042	0.5406	0.5304	0.5337
PROPHET		0.5014	0.7687	1.1886	2.6113
NN		0.5238	0.6114	0.5526	0.5303
SVR		0.5410	0.5489	0.5234	0.5699
RF		0.5437	0.5433	0.5813	0.5699
XGBOOST		0.5607	0.5647	0.5145	0.4901
LGBM		0.5302	0.5108	0.4914	0.4915
SARIMA		6	0.4888	/	/
RIDGE	0.5004		0.6592	0.5557	0.5962
LASSO	0.5188		0.5664	0.5319	0.5329
ENET	0.5097		0.5642	0.5392	0.5489
PROPHET	0.5325		0.8255	1.7516	2.8295
NN	0.5089		0.7542	0.5894	0.6601
SVR	0.5385		0.5671	0.5453	0.6001
RF	0.5547		0.5620	0.6299	0.6001
XGBOOST	0.5500		0.5839	0.5171	0.5688
LGBM	0.5431		0.5173	0.5496	0.5139
SARIMA	9		0.4864	/	/
RIDGE		0.4904	0.7095	0.5373	0.6251
LASSO		0.5039	0.5089	0.5346	0.5191
ENET		0.5033	0.5176	0.5346	0.5214
PROPHET		0.5147	1.0221	1.2750	5.9239
NN		0.5036	0.8434	0.5597	0.6644
SVR		0.5377	0.5460	0.5469	0.5790
RF		0.5224	0.5496	0.6103	0.5790
XGBOOST		0.5292	0.5434	0.5313	0.5322
LGBM		0.5292	0.4981	0.4949	0.5080

Source: Authors' calculations.

Accordingly, two models are selected for the final inflation forecasting exercise: SARIMA, as the traditional time-series model that consistently delivers the strongest results, and LightGBM, which has proven to be the most successful among ML approaches. Using these models, monthly inflation rates (month-over-month) are forecast for each of the 14 horizons. The resulting monthly forecasts are then transformed, as described in section 3.4, into annual inflation rates (year-over-year) by incorporating the contributions of realised monthly inflation rates. This enables comparison with the European Commission's forecasts across the three forecasting cycles. Figure 2 presents the comparison of actual average annual inflation with the European Commission's projections and the forecasts generated by the SARIMA and LightGBM models for all three forecasting scenarios: (1) the autumn forecast for the current year ($h = 2$); (2) the spring forecast for the current year ($h = 8$); and (3) the autumn forecast for the following year ($h = 14$).

FIGURE 2

Forecasted average annual inflation rate by the European Commission compared with forecasts from the SARIMA and LightGBM models (in %)



Note: Owing to the longer forecasting horizon of 14 months, the first year for which a complete annual inflation forecast can be computed in the last panel is 2018, whereas for shorter horizons (2 and 8 months) forecasts begin in 2017. Since the most recent publication of the European Commission available at the time of writing was released in autumn 2024, the final estimate for the current-year forecast refers to 2024. In the case of the next-year forecast (the 14-month horizon), the most recent projection refers to 2025.

Source: Authors' calculations; European Commission.

Figure 2 shows that the SARIMA and LightGBM models outperform the European Commission's forecasts in all forecasting scenarios. For the shortest horizon ($h = 2$, autumn forecast for the current year), the differences between the models and the European Commission are small, as most observations for that year are already known; nevertheless, our models provide slightly more accurate estimates. At the medium horizon ($h = 8$, spring forecast for the current year), the deviations between forecasted and realised values become larger, which is expected given the greater uncertainty and the longer time span until the end of the year. Even so, SARIMA and LightGBM continue to exhibit smaller errors relative to actual outcomes than the European Commission. The largest deviations occur at the longest horizon ($h = 14$, autumn forecast for the following year), clearly illustrating the difficulty of predicting inflation more than a year in advance. Even in this case, the forecasts generated by the two models remain closer to the realised values than those of the European Commission, thereby confirming their superiority under conditions of pronounced uncertainty. Finally, the figure reveals a clear distinction between SARIMA and LightGBM. SARIMA forecasts inflation more accurately during the stable pre-COVID period, whereas LightGBM delivers more precise predictions in the volatile post-pandemic environment.

5 DISCUSSION

The central question addressed in this study is whether ML models outperform benchmark models in forecasting inflation in Croatia and, if so, which ML approach performs best. The results suggest that ML methods achieve forecasting accuracy comparable to, and in some cases exceeding, that of standard univariate econometric models. This finding is consistent with recent evidence showing that ML techniques can improve inflation forecasts relative to traditional approaches (Medeiros et al., 2021; Araujo and Gaglianone, 2023). Among the ML methods considered, tree-based models perform particularly well, in line with results reported in related empirical studies.

Beyond model choice, forecasting performance is strongly influenced by the quantity and informational content of the predictors. Recent studies emphasize the benefits of data-rich environments, where models exploit large sets of potential predictors (e.g. Medeiros et al., 2021; Araujo and Gaglianone, 2023). In contrast, this paper adopts a parsimonious, *ad hoc* selection of key variables. While this simplifies the modelling framework, it may limit the gains typically associated with ML in data-rich settings. Nevertheless, given the available sample size, the selected predictors appear adequate: LightGBM performs competitively relative to SARIMA, suggesting that the essential information for forecasting inflation is captured by this reduced feature set. This approach is also motivated by structural constraints of the Croatian data environment, characterized by relatively short time series and a limited number of high-frequency macroeconomic indicators, a limitation also noted for other small economies (e.g. Ivaşcu, 2023). As longer time series become available, the relative performance of ML models, particularly tree-based methods, may improve.

Differences between our results and those reported for larger economies can also be explained by specific features of the Croatian macroeconomic environment. For much of the sample period, monetary policy relied on the exchange rate as the nominal anchor, while external shocks played a dominant role in shaping inflation dynamics (Globan, Arčabić and Sorić, 2015). Under such conditions, inflation exhibits strong persistence and pronounced trend and seasonal components, making it particularly amenable to models such as SARIMA. Previous studies document the strong performance of ARIMA-type models in forecasting Croatian inflation (Pufnik and Kunovac, 2006) and in explaining its dynamics (Živko and Bošnjak, 2017), while our results confirm that SARIMA remains a reliable benchmark in this setting. The pronounced autoregressive nature of inflation implies that models explicitly exploiting its temporal structure can be highly effective, helping to explain the strong performance of SARIMA relative to more complex alternatives.

Finally, the results indicate that, over the sample period, our models consistently produced forecasts closer to realised inflation outcomes than those published by the European Commission, even in periods of heightened uncertainty. These results should, however, be interpreted with caution. Forecasts produced by institutions such as the European Commission may incorporate elements of forward guidance aimed at shaping public expectations, which can generate systematic deviations from realised outcomes, particularly during episodes of large shocks. While such practices may support short-term expectation anchoring, persistent forecast errors may ultimately weaken institutional credibility, raising the question of whether expectation management through official forecasts enhances macroeconomic stability or undermines trust over time. This trade-off remains an open empirical question with important implications for the design of official forecasting frameworks.

6 CAVEATS

This section outlines several important limitations related to data availability and methodological choices underlying the empirical analysis.

A first limitation concerns the use of single-extraction Google Trends series. Recent studies have shown that Google Trends data are subject to sampling variation, implying that repeated extractions of the same query may yield different values and thereby raising concerns regarding reproducibility (e.g. Cebrián and Domenech, 2024). To address this issue, some contributions propose averaging multiple extractions of the same query. However, implementing such procedures would substantially increase the data-collection burden and delay model estimation. Consequently, the analysis relies on a single extraction per query, a practice that has also been adopted in earlier studies (e.g. Choi and Varian, 2012). Nevertheless, this choice represents a limitation that should be considered when interpreting the results.

A second limitation relates to hyperparameter optimisation. Due to computational constraints, the range of hyperparameters considered in the grid search was necessarily restricted, although it was not excessively narrow. Future work could address this limitation by expanding the grid search to cover a wider set of hyperparameter values, provided sufficient computational resources are available. Alternatively, the use of alternative hyperparameter tuning strategies beyond grid search, such as random search or Bayesian optimisation, could be considered in future research.

A third limitation concerns differences in model selection and evaluation procedures across model classes. The SARIMA models were specified using the Akaike information criterion (AIC), in line with standard practice in the econometric time-series literature, where information criteria are commonly employed for model selection. In contrast, machine learning models were trained and evaluated using an expanding-window cross-validation approach, which is more typical in the contemporary forecasting literature. This design choice was motivated by the intention to place each modelling approach in its natural methodological setting. Nevertheless, the use of different training and evaluation frameworks may limit the strict comparability of results across model classes. Future research could address this issue by applying a unified evaluation strategy (such as CV) to all models.

7 CONCLUSION

After several years of stability and low inflation in advanced economies, inflation has once again moved to the forefront of attention in both academic and policy circles. The sudden shocks associated with the COVID-19 pandemic, disruptions in global supply chains and increased geopolitical tensions triggered a sharp rise in prices, posing a significant challenge for policymakers, particularly central banks, which rely on accurate inflation forecasts to design effective monetary measures. In this context, timely and precise inflation projections are essential. This paper therefore examines the potential of modern machine learning methods for forecasting inflation in Croatia, a country where the literature on inflation forecasting, and especially on the application of machine learning, remains notably limited.

The findings show that no single model performs best across all time periods and feature sets. However, SARIMA, a traditional time-series model, stands out as the most reliable and consistently accurate model overall. Tree-based models, especially LightGBM, also exhibit strong predictive capabilities and outperform other machine learning methods across a variety of feature combinations. As data availability expands and new macroeconomic conditions are incorporated over time, it is reasonable to expect that models such as LightGBM will increasingly assume a leading role in inflation forecasting.

In the final comparison with the European Commission's projections, both SARIMA and LightGBM produce more accurate forecasts of average annual inflation across all forecasting scenarios, even at longer horizons. This further

confirms their predictive superiority. The results also reveal an important pattern: before the COVID-19 crisis, during a period of greater macroeconomic stability, SARIMA generated more accurate forecasts than LightGBM, whereas in the post-COVID period LightGBM delivered more precise predictions than SARIMA.

Overall, the results point to a considerable potential for machine learning methods in inflation forecasting. By incorporating diverse input variables and evaluating models across different economic regimes, this study contributes to the empirical literature and highlights several avenues for further research. Given that inflation is a multidimensional and complex phenomenon, sophisticated tools are required, and machine learning models, when applied with due care, appear to offer a promising solution.

Disclosure statement

The authors have no conflicts of interest to declare.

REFERENCES

1. Ang, A., Bekaert, G. and Wei, M., 2005. Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4), pp. 1163-1212. <https://doi.org/10.1016/j.jmoneco.2006.04.006>
2. 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. <https://doi.org/10.1016/j.lacb.2023.100087>
3. Atkeson, A. and Ohanian, L. E., 2001. Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), pp. 2-11. <https://doi.org/10.21034/qv.2511>
4. Barhoumi, K., Darné, O. and Ferrara, L., 2009. Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, 29(1-2), pp. 132-144. <https://doi.org/10.1002/for.1162>
5. Blinder, A. S., 1997. Is there a core of practical macroeconomics that we should all believe? *The American Economic Review*, 87(2), pp. 240-243.
6. Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24(2), pp. 123-140. <https://doi.org/10.1007/BF00058655>
7. Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5-32. <https://doi.org/10.1023/A:1010933404324>
8. Cebrián, E. and Domenech, J., 2024. Addressing Google Trends inconsistencies. *Technological Forecasting and Social Change*, 202, 123318. <https://doi.org/10.1016/j.techfore.2024.123318>
9. 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*. New York: Association for Computing Machinery, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
10. Chen, Y.-C., Turnovsky, S. J. and Zivot, E., 2014. Forecasting inflation using commodity price aggregates. *Journal of Econometrics*, 183(1), pp. 117-134. <https://doi.org/10.1016/j.jeconom.2014.06.013>
11. Choi, H. and Varian, H., 2012. Predicting the present with Google Trends. *Economic Record*, 88(s1), pp. 2-9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
12. Drucker, H. [et al.], 1997. Support vector regression machines. *Advances in Neural Information Processing Systems*, 28, pp. 779-784.
13. Eickmeier, S. and Ziegler, C., 2008. How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting*, 27(3), pp. 237-265. <https://doi.org/10.1002/for.1056>
14. Forni, M. [et al.], 2003. Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, 50(6), pp. 1243-1255. [https://doi.org/10.1016/s0304-3932\(03\)00079-5](https://doi.org/10.1016/s0304-3932(03)00079-5)

15. Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), pp. 1189-1232. <https://doi.org/10.1214/aos/1013203451>
16. Garcia, M. G. P., Medeiros, M. C. and Vasconcelos, G. F. R., 2017. Real-time inflation forecasting with high-dimensional models: the case of Brazil. *International Journal of Forecasting*, 33(3), pp. 679-693. <https://doi.org/10.1016/j.ijforecast.2017.02.002>
17. Globan, T., Arčabić, V. and Sorić, P., 2015. Inflation in new EU member states: a domestically or externally driven phenomenon? *Emerging Markets Finance and Trade*, 52(1), pp. 154-168. <https://doi.org/10.1080/1540496x.2014.998547>
18. Groen, J. J. J., Paap, R. and Ravazzolo, F., 2013. Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics*, 31(1), pp. 29-44. <https://doi.org/10.1080/07350015.2012.727718>
19. Hoerl, A. E. and Kennard, R. W., 1970. Ridge regression: applications to non-orthogonal problems. *Technometrics*, 12(1), pp. 69-82. <https://doi.org/10.2307/1267351>
20. Hyndman, R. J. and Athanasopoulos, G., 2021. *Forecasting: principles and practice*. 3rd ed. Melbourne: OTexts.
21. Hyndman, R. J. and Koehler, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), pp. 679-688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
22. Ivaşcu, C., 2023. Can machine learning models predict inflation? *Proceedings of the International Conference on Business Excellence*, 17(1), pp. 1748-1756. <https://doi.org/10.2478/picbe-2023-0155>
23. Ke, G. [et al.], 2017. LightGBM: a highly efficient gradient boosting decision tree. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Red Hook, NY: Curran Associates Inc., pp. 3149-3157.
24. Kunovac, D., 2007. Factor model forecasting of inflation in Croatia. *Financial Theory and Practice*, 31(4), pp. 371-393.
25. Masini, R. P., Medeiros, M. C. and Mendes, E. F., 2021. Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1), pp. 76-111. <https://doi.org/10.1111/joes.12429>
26. Medeiros, M. C. [et al.], 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business and Economic Statistics*, 39(1), pp. 98-119. <https://doi.org/10.1080/07350015.2019.1637745>
27. Medeiros, M. C. and Mendes, E. F., 2016. ℓ_1 -regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, 191(1), pp. 255-271. doi.org/10.1016/j.jeconom.2015.10.011
28. Monteforte, L. and Moretti, G., 2013. Real-time forecasts of inflation: the role of financial variables. *Journal of Forecasting*, 32(1), pp. 51-61. <https://doi.org/10.1002/for.1250>

29. Naghi, A. A., O'Neill, E. and Zaharieva, M. D., 2024. The benefits of forecasting inflation with machine learning: new evidence. *Journal of Applied Econometrics*, 39(7), pp. 1321-1331. <https://doi.org/10.1002/jae.3088>
30. Nakamura, E., 2004. Inflation forecasting using a neural network. *Economics Letters*, 86(3), pp. 373-378. <https://doi.org/10.1016/j.econlet.2004.09.003>
31. Payne, J. E., 2002. Inflationary dynamics of a transition economy: the Croatian experience. *Journal of Policy Modeling*, 24(3), pp. 219-230. [https://doi.org/10.1016/s0161-8938\(02\)00106-0](https://doi.org/10.1016/s0161-8938(02)00106-0)
32. Pufnik, A. and Kunovac, D., 2006. Short-term inflation forecasting in Croatia using seasonal ARIMA processes. *Istraživanja* 1-18.
33. Rumelhart, D. E., Hinton, G. E. and Williams, R. J., 1986. Learning representations by back-propagating errors. *Nature*, 323(6088), pp. 533-536. <https://doi.org/10.1038/323533a0>
34. Scornet, E., Biau, G. and Vert, J.-P., 2015. Consistency of random forests. *The Annals of Statistics*, 43(4), pp. 1716-1741. <https://doi.org/10.1214/15-aos1321>
35. Sorić, P. and Lolić, I., 2017. Economic uncertainty and its impact on the Croatian economy. *Public Sector Economics*, 41(4), pp. 443-477. <https://doi.org/10.3326/pse.41.4.3>
36. Stock, J. H. and Watson, M., 2003. Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature*, 41(3), pp. 788-829. <https://doi.org/10.1257/002205103322436197>
37. Taylor, S. J. and Letham, B., 2018. Forecasting at scale. *The American Statistician*, 72(1), pp. 37-45. <https://doi.org/10.1080/00031305.2017.1380080>
38. Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58(1), pp. 267-288.
39. Ülke, V., Sahin, A. and Subasi, A., 2018. A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing and Applications*, 30(5), pp. 1519-1527. <https://doi.org/10.1007/s00521-016-2766-x>
40. 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), pp. 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
41. Živko, I. and Bošnjak, M., 2017. Time series modeling of inflation and its volatility in Croatia. *Notitia – časopis za ekonomske, poslovne i društvene teme*, 3(1), pp. 1-10. <https://doi.org/10.32676/n.3.1>

CONVERSION TO YEAR-OVER-YEAR INFLATION RATES

Table A1 illustrates the procedure for converting monthly inflation rates into year-over-year inflation rates, which facilitates understanding of equations (A1) to (A4). The year-over-year inflation rate in month t is defined as:

$$1 + \pi_t^{vy} = \frac{P_t}{P_{t-12}} = \frac{P_t}{P_{t-h}} \cdot \frac{P_{t-h}}{P_{t-12}} \quad (\text{A1})$$

where P_t denotes the consumer price index in month t . The contribution of monthly inflation rates to the year-over-year rate for the most recent h months is given by:

$$\frac{P_t}{P_{t-h}} = \prod_{s=t-h+1}^t (1 + \hat{\pi}_s^{mom}) \quad (\text{A2})$$

In equation (A2), we use the forecasted monthly inflation rates $\hat{\pi}_s^{mom}$, while in equation (A3) we retain the realised monthly rates π_s^{mom} . The contribution of realised monthly changes in the HICP index over the months that remain in the base after the most recent h months drop out is expressed as:

$$\frac{P_{t-h}}{P_{t-12}} = \prod_{s=t-11}^{t-h} (1 + \pi_s^{mom}) \quad (\text{A3})$$

Combining equations (A2) and (A3) yields:

$$\hat{\pi}_t^{vy} = \left[\prod_{s=t-h+1}^t (1 + \hat{\pi}_s^{mom}) \right] \left[\prod_{s=t-11}^{t-h} (1 + \pi_s^{mom}) \right] - 1 \quad (\text{A4})$$

TABLE A1

RMSE expressed as indices relative to the SARIMA model for the pre-COVID-19 period

Model	Horizon	lag	unconv	conv	all	
SARIMA	3	100.0	/	/	/	
RIDGE		109.9	108.4	111.2	106.5	
LASSO		113.9	118.0	114.0	111.7	
ENET		114.3	117.2	113.9	111.9	
PROPHET		119.3	215.6	567.4	304.8	
NN		110.1	131.1	111.1	116.8	
SVR		116.3	120.2	121.7	123.2	
RF		122.7	127.2	131.7	123.2	
XGBOOST		116.3	107.7	125.1	103.3	
LGBM		108.1	117.9	111.7	109.2	
SARIMA		6	100.0	/	/	/
RIDGE			112.6	109.4	113.2	108.7
LASSO			118.4	129.6	125.1	126.5
ENET			121.7	129.7	126.6	125.1
PROPHET	131.5		218.6	803.2	368.1	
NN	112.7		123.3	138.9	126.1	
SVR	117.7		122.8	121.3	141.4	
RF	118.6		131.8	161.1	141.4	
XGBOOST	104.1		108.3	111.7	113.8	
LGBM	94.9		119.3	134.6	108.9	

Model	Horizon	lag	unconv	conv	all
SARIMA	9	100.0	/	/	/
RIDGE		109.0	106.6	108.2	103.5
LASSO		114.1	115.6	115.5	116.0
ENET		115.3	118.1	115.8	116.1
PROPHET		122.3	249.5	523.9	273.0
NN		105.7	111.8	111.0	114.9
SVR		117.2	111.0	111.4	123.4
RF		105.6	110.1	152.5	123.4
XGBOOST		90.9	102.7	99.6	101.4
LGBM		102.0	105.6	90.6	96.5

Source: Authors' calculations.

TABLE A2

RMSE expressed as indices relative to the SARIMA model for the post-COVID-19 period

Model	Horizon	lag	unconv	conv	all
SARIMA	3	100.0	/	/	/
RIDGE		107.6	118.7	116.4	112.5
LASSO		105.3	112.0	107.3	109.8
ENET		103.8	112.2	109.8	110.9
PROPHET		102.1	149.2	160.8	577.7
NN		109.4	127.3	116.3	109.9
SVR		112.6	113.8	107.2	118.5
RF		112.1	111.1	119.7	118.5
XGBOOST		117.4	119.8	104.3	102.4
LGBM		111.3	104.8	101.1	101.6
SARIMA		6	100.0	/	/
RIDGE	100.4		138.1	113.5	123.6
LASSO	102.8		112.1	104.9	104.7
ENET	101.0		111.5	106.3	108.9
PROPHET	104.7		159.3	218.2	603.6
NN	102.5		158.2	117.2	136.1
SVR	108.7		114.6	109.7	119.3
RF	112.4		111.8	122.8	119.3
XGBOOST	113.5		120.8	104.6	116.5
LGBM	113.1		103.3	108.3	104.3
SARIMA	9		100.0	/	/
RIDGE		99.2	150.9	110.2	131.3
LASSO		101.6	102.5	108.3	104.5
ENET		101.2	104.1	108.3	105.1
PROPHET		102.7	202.6	188.8	1298.5
NN		103.0	181.0	115.0	138.9
SVR		109.2	112.2	111.9	117.6
RF		107.4	113.2	119.9	117.6
XGBOOST		111.2	112.8	110.0	109.9
LGBM		109.6	101.6	102.8	105.0

Source: Authors' calculations.

TABLE A3

RMSE expressed as indices relative to the SARIMA model for the full observation period

Model	Horizon	lag	unconv	conv	all	
SARIMA	3	100.0	/	/	/	
RIDGE		108.2	117.5	115.9	111.9	
LASSO		107.4	113.0	108.8	110.7	
ENET		105.7	113.3	111.1	111.8	
PROPHET		105.1	161.1	249.1	547.2	
NN		109.8	128.1	115.8	111.1	
SVR		113.4	115.0	109.7	119.4	
RF		113.9	113.9	121.8	119.4	
XGBOOST		117.5	118.3	107.8	102.7	
LGBM		111.1	107.0	103.0	103.0	
SARIMA		6	100.0	/	/	/
RIDGE			102.4	134.9	113.7	122.0
LASSO	106.1		115.9	108.8	109.0	
ENET	104.3		115.4	110.3	112.3	
PROPHET	108.9		168.9	358.3	578.9	
NN	104.1		154.3	120.6	135.0	
SVR	110.2		116.0	111.6	122.8	
RF	113.5		115.0	128.9	122.8	
XGBOOST	112.5		119.5	105.8	116.4	
LGBM	111.1		105.8	112.4	105.1	
SARIMA	9		100.0	/	/	/
RIDGE			100.8	145.9	110.5	128.5
LASSO		103.6	104.6	109.9	106.7	
ENET		103.5	106.4	109.9	107.2	
PROPHET		105.8	210.1	262.1	1217.9	
NN		103.5	173.4	115.1	136.6	
SVR		110.5	112.3	112.4	119.0	
RF		107.4	113.0	125.5	119.0	
XGBOOST		108.8	111.7	109.2	109.4	
LGBM		108.8	102.4	101.7	104.4	

Source: Authors' calculations.

TABLE A4

Average RMSE by feature set used

Period	Variables	SARIMA	RIDGE	LASSO	ENET	PROPHET	NN	SVR	RF	XGB	LGBM
Before the pandemic	<i>lag</i>	0.291	0.321	0.336	0.340	0.361	0.318	0.340	0.336	0.302	0.296
	<i>unconv</i>	/	0.314	0.352	0.354	0.662	0.355	0.343	0.358	0.309	0.332
	<i>conv</i>	/	0.322	0.343	0.345	1.833	0.349	0.343	0.431	0.326	0.326
	<i>all</i>	/	0.309	0.343	0.342	0.916	0.347	0.376	0.376	0.308	0.305
Average RMSE		0.291	0.317	0.343	0.345	0.943	0.342	0.351	0.375	0.311	0.315
	<i>lag</i>	0.576	0.59	0.595	0.588	0.594	0.604	0.634	0.637	0.657	0.642
	<i>unconv</i>	/	0.784	0.627	0.629	0.982	0.897	0.654	0.645	0.679	0.595
	<i>conv</i>	/	0.653	0.615	0.623	1.092	0.669	0.631	0.696	0.612	0.600
Average RMSE		0.576	0.683	0.612	0.616	4.770	0.740	0.682	0.682	0.632	0.597
	<i>lag</i>	0.484	0.502	0.512	0.506	1.860	0.728	0.651	0.665	0.645	0.608
	<i>unconv</i>	/	0.643	0.538	0.541	0.516	0.512	0.539	0.540	0.547	0.534
	<i>conv</i>	/	0.549	0.529	0.535	1.405	0.567	0.554	0.552	0.564	0.509
Average RMSE		0.484	0.585	0.527	0.535	3.788	0.618	0.583	0.583	0.530	0.504
	<i>lag</i>	0.484	0.570	0.526	0.529	1.645	0.608	0.554	0.571	0.540	0.515
	<i>unconv</i>	/	0.585	0.527	0.535	1.405	0.567	0.539	0.607	0.521	0.512
	<i>conv</i>	/	0.585	0.527	0.535	3.788	0.618	0.583	0.583	0.530	0.504

Source: Authors' calculations.