

JEL Classification: E37, E31, C53

Keywords: machine learning, forecasting, inflation, Slovakia, Lasso, Boosting

# Forecasting Inflation in Slovakia Using Machine Learning

Ádám CSÁPAI - Bratislava University of Economics and Business, Faculty of Economic Informatics, Department of Statistics, and Institute of Economic Research, Slovak Academy of Sciences (adam.csapai@euba.sk) corresponding author

## Abstract

*Reliable inflation forecasts are essential for effective policy interventions and maintaining economic stability. This paper evaluates the use of machine learning methods to forecast inflation in Slovakia over different time frames, including the high volatility period of the COVID-19 pandemic and its aftermath. We work with a comprehensive dataset from the National Bank of Slovakia containing 16 years of observations of 380 variables. We find that methods such as Lasso and Boosting, which utilize regularization and address nonlinearities, outperform traditional models in terms of predictive power. By deploying nonlinear approaches on a regularized data set, we are able to enhance these positive outcomes further. The implementation of these advanced methods allows policymakers to improve their decision-making capabilities and prepare more accurate inflation forecasts.*

## 1. Introduction

Forecasting plays a vital role in assessing the economic condition and guiding economic policymaking. It is foundational for crafting government budgets and crucial for policymakers, like Central Banks, to time interventions effectively based on forecasts of key economic indicators such as GDP, inflation, and unemployment. Current models, however, often fail to capture the true dynamics between economic variables. For example, Medeiros et al. (2021) illustrate how governments and international bodies, especially the ECB, tend to consistently overestimate inflation projections. Such discrepancies can lead to significant welfare losses and skew inflation expectations, underlining the need for more precise forecasting models. What is more, Medeiros et al. (2021) also highlights that the extensive literature on inflation forecasting reveals that Phillips curve-based models often fail to produce accurate inflation predictions. While Stock and Watson (1999) identifies numerous production-related variables as potential predictors of US inflation, Atkeson and Ohanian (2001)

---

<https://doi.org/10.32065/CJEF.2025.04.03>

The financial support was provided by the Scientific grant agency (VEGA) of the Slovak Ministry of Education, Research, Development and Youth under the "VEGA No. 1/0047/23 The importance of spatial spillover effects in the context of the EU's greener and carbon-free Europe priority." project and the Early Stage Grant number "A-25-103/3020-11 Macroeconomic forecasting using advanced machine learning" of the University of Economics in Bratislava. I would also like to thank the anonymous reviewers for their valuable comments and suggestions.

demonstrate that the Phillips curve frequently underperforms compared to even simple naive models. This has led researchers to explore alternative models and variables to enhance inflation forecasts. Among the explored variables are financial indicators (Forni et al., 2003), commodity prices (Chen et al., 2014), and expectation variables (Groen et al., 2013). However, there is no consistent evidence that these models consistently surpass the established benchmarks.

With the emergence of big data, enhanced computational capabilities, and advances in statistical learning, economists now have access to a variety of new methods, including those based on machine learning. These methods have become increasingly popular in macroeconomic applications over the last decade, particularly in the last five to six years.

This paper presents a comparative analysis of machine learning models for forecasting inflation in Slovakia. We specifically forecast inflation as measured by the harmonized index of consumer prices (HICP) over five time horizons: 1, 3, 6, 9, and 12 months ahead. We aim to determine the relative efficacy of machine learning models in capturing the complexities of the Slovak economy compared to conventional methods.

*Our contributions to macroeconomic forecasting include several noteworthy approaches. Firstly, this study is the first to utilize advanced machine learning techniques to analyze the nonlinear relationships between Slovakian inflation and its predictors. Secondly, we implement regularization to streamline the model inputs, which helps improve performance against standard benchmarks and, as stated in the literature review, is not affected by the drawbacks of factor models. To our knowledge, there is no similar study focused on Slovakia or, indeed, any small open economy in a monetary union. The existing studies mostly focus on large developed economies like the US. Thirdly, we incorporate data from the COVID-19 pandemic to evaluate how well these models perform under conditions of heightened economic volatility and uncertainty. Lastly, we employ a mostly unexplored approach similar to that of Medeiros et al. (2021) by integrating regularization with nonlinear models enhancing the accuracy of Random Forest models through careful variable selection.*

Our contributions to macroeconomic forecasting include several noteworthy approaches. Firstly, this study is, to our knowledge, the first to apply a suite of advanced machine learning techniques (Lasso, Boosting, Random Forest, and a hybrid Lasso–Random Forest approach) to inflation forecasting for Slovakia, thereby extending the literature to a small open economy within the euro area. Secondly, we implement regularization to streamline the model inputs, which helps improve performance against standard benchmarks and, as stated in the literature review, is not affected by the drawbacks of factor models. Thirdly, we explicitly evaluate model performance during the COVID-19 pandemic, allowing us to examine the robustness of ML methods in a period of extreme macroeconomic volatility. Lastly, we employ a mostly unexplored approach similar to that of Medeiros et al. (2021) by integrating regularization with nonlinear models, demonstrating that this combination can further enhance forecasting accuracy in small-sample, high-dimensional settings. Together, these contributions position our paper as both an extension of prior research and a country-specific case study with direct relevance.

The paper is organized as follows. The second section reviews the existing literature on macroeconomic forecasting with machine learning methods. The third

section describes our data and methodology. The fourth section presents our results while the fifth section concludes.

## 2. Literature Review

Vargas (2020) observes that advancements in computing power, statistical learning theory, and the increased availability of big data over the past decade have led to the adoption of machine learning methods in economic forecasting. Historically, these methods were primarily applied to classification tasks, such as predicting loan delinquencies or consumer purchasing decisions, where they surpassed the performance of traditional models. Although these applications focus on discrete variables, Vargas (2020) points out that machine learning can also effectively predict continuous variables such as GDP or inflation. According to Vargas (2020), machine learning's ability to capture nonlinear relationships in the data enhances our understanding of complex economic phenomena like asymmetric business cycles, stock market volatility, and regime switches.

The conclusions of Vargas (2020) are supported by Athey (2019), who notes that over the past decade, economists have increasingly turned to machine learning for forecasting, especially when large datasets are available. These methods are mostly used for forecasting financial variables such as stock prices, as there is ample data. An example case study of machine learning application is provided by Nechvátalová (2024), who uses machine learning methods to forecast global expected returns of stocks across different horizons. Her results reveal that return predictability diminishes as the forecasting horizon extends, both in the U.S. and globally. However, she demonstrates that firm-specific characteristics can still yield profitable insights, even after factoring in transaction costs, particularly for longer forecasting periods.

While Nechvátalová (2024) focuses on stock prices, Athey (2019) primarily discusses the use of microdata for forecasting, and Mullainathan and Spiess (2017) investigate the application of these methods in macroeconomic forecasting. All of these authors argue that macroeconomists should add machine learning tools to their forecasting arsenal.

Bolhuis and Rayner (2020) analyze the theoretical aspects of Ordinary Least Squares (OLS) in comparison to factor models and machine learning. They point out that OLS predictions may be unsuitable due to issues like predictor relevance, nonlinearity, collinearity, and high dimensionality. While factor models have been central in data-driven forecasting for decades and can mitigate problems related to dimensionality and collinearity, they fall short when it comes to addressing nonlinearity and predictor relevance, often resulting in less accurate forecasts. Bolhuis and Rayner (2020) state that machine learning methods, in contrast, excel at capturing nonlinear dynamics within high-dimensional datasets. They adeptly learn from complex historical data relationships without making unwarranted future projections. Thus, despite the effectiveness of simple data-driven models, their drawbacks have prompted researchers to explore machine learning as a viable alternative for macroeconomic forecasting.

To elaborate further, we describe these methods within a macroeconomic framework. Masini et al. (2021) review recent progress in the field and note that machine learning encompasses various approaches. Specifically for macroeconomic

forecasting, they describe machine learning as the integration of automated algorithms and statistical methods to identify patterns in large datasets. They also summarize the differences between the three widely used categories of machine learning, namely supervised, unsupervised, and reinforcement learning.

Like Masini et al. (2021), we utilize supervised learning techniques for our forecasting purposes. Supervised learning involves training the model on data that is structured as input-output pairs. For example, in linear regression models, the inputs (or predictors) and the outputs (target variable forecasts) are paired clearly. The model learns a function that maps the input to the output, aiming to predict the value of the target variable effectively. The specifics of individual machine learning methods employed are discussed in the methodological section of this paper.

The empirical evidence on the macroeconomic forecasting capacity of machine learning methods is scarce but compelling. Chakraborty and Joseph (2017) apply machine learning to predict medium-term inflation rates and report that their selected methods can outperform traditional benchmarks by as much as 29%. Similarly, Jung et al. (2018) employ methods like Elastic Net, SuperLearner, and Recurrent Neural Networks to forecast macroeconomic variables for seven advanced and emerging economies, achieving results that surpass the benchmark World Economic Outlook (WEO) forecasts.

Coulombe et al. (2022) conduct a forecasting exercise using US data from the past four decades to explore the factors contributing to the superior performance of machine learning methods. They forecast multiple target variables across five different time horizons using various models, pinpointing the critical success factors for machine learning in forecasting. Firstly, they find that nonlinearities significantly impact the data generating process in the US, enhancing the predictive power of machine learning models. Secondly, they demonstrate that regularization techniques greatly improve forecasting accuracy, outperforming traditional factor models. Lastly, they establish that  $k$ -fold cross-validation is more effective for model selection than the traditional information criteria typically used in econometrics.

Medeiros et al. (2021) delve into why regularization and nonlinearities are critical in macroeconomic forecasting. Starting with regularization: in recent years, as big data becomes more prevalent, macroeconomists are reevaluating the significance of each variable within these large datasets. In addition to factor models, there has been an increased use of shrinkage and variable selection techniques. Notably, Medeiros and Mendes (2016), and Giannone et al. (2021) recommend Lasso models for macroeconomic forecasting, which have shown to outperform traditional benchmarks. These studies primarily focus on US inflation data and short-term forecasting. The most extensive study by Medeiros et al. (2021) reveals that although shrinkage methods are slightly less precise than tree-based methods in the US, they significantly enhance forecasting accuracy.

Regarding nonlinearities, Medeiros et al. (2021) also provide a theoretical backdrop. For example, the relationship between inflation and employment is nonlinear and depends on economic slackness. Additionally, nonlinearity arises from uncertainty; for instance, the nonlinear hiring practices due to the high costs associated with firing employees. Furthermore, the zero lower bound (ZLB) on nominal interest rates introduces nonlinearity between inflation, employment, and interest rates, especially under unconventional monetary policies. The role of houses as collateral

and their interaction with monetary policy and financial intermediation further complicates this nonlinearity. As evidenced during the Great Recession, the burst of a housing bubble can precipitate severe credit downturns, marked by nonlinear dynamics.

Our contribution builds on these findings by explicitly testing whether similar benefits from regularization and nonlinear modeling hold in a small open economy, using a comprehensive Slovak dataset covering the Great Recession, a prolonged near-ZLB period, and the COVID-19 pandemic.

In this section we provide theoretical and empirical justification for the employment of machine learning methods in macroeconomic forecasting. In the next chapter we describe our dataset, the forecasting setup and the methods selected for forecasting.

### 3. Methodology and Methods

#### 3.1 Data and Forecasting Setup

We assess the forecasting capability of machine learning methods compared to a benchmark, focusing on inflation measured by the harmonized index of consumer prices (HICP). Forecasts are generated at horizons of 1, 3, 6, 9, and 12 months ahead.

**Database.** The data are obtained from the National Bank of Slovakia (NBS) database.<sup>1</sup>

The database is organized into eight categories: GDP, labor market, prices, government, balance of payments, external environment, monthly “soft” indicators, and the financial market. All data are seasonally adjusted by NBS when necessary. We construct two datasets. The first excludes the monthly “soft” indicators and comprises 274 variables with 188 monthly observations. Variables with missing values or that do not cover our full period are left out from the original NBS dataset. The second dataset consists only of the 104 soft indicators, also with 188 observations. The full variable list is provided in Appendix A.

Soft indicators are considered separately because of their timeliness. Coulombe et al. (2022) argue that these variables, available earlier than official macroeconomic releases, may provide useful signals in real time. Evaluating them separately allows us to assess their contribution during periods requiring rapid policy responses, such as the COVID-19 pandemic. Both datasets cover January 2008 to December 2023.

We use only ex post available data and do not attempt to replicate publication lags. This follows standard practice in the machine learning forecasting literature (Maehashi and Shintani, 2020; Masini et al., 2021). Predictors are not pre-selected. Their relevance is determined endogenously by the models. Appendix A documents all available predictors (the time series we use) and their transformations.

**Preprocessing.** Before model training, all model inputs from Appendix A are standardized to zero mean and unit variance. Where necessary, log transformations and differencing are applied to achieve stationarity, as indicated in the appendix. These steps follow Zhang and Qi (2005), who show that machine learning models perform best with standardized inputs.

---

<sup>1</sup><https://nbs.sk/statisticke-udaje/vybrane-makroekonomicke-ukazovatele/makroekonomicka-databaza/>

**Rolling window design.** Forecasts are produced using a 72-month (six-year) rolling window. At time  $t = R$ , the model is estimated on data  $[R - 71, \dots, R]$  and a direct  $h$ -step-ahead forecast  $\hat{y}_{R+h}$  is produced. The window then shifts forward by one observation, and the process repeats. All parameters and hyperparameters are re-estimated in each iteration.

The procedure is implemented separately for each forecast horizon  $h = (1, 3, 6, 9, 12)$ . At each step, direct multi-step forecasts are generated for periods  $[R + 1, \dots, R + h]$ , ensuring that evaluation is truly out-of-sample and does not rely on iterated one-step predictions. This is implemented by shifting the target variable accordingly in each rolling window. For every iteration, forecast errors and mean squared errors are computed and stored by horizon. These are later aggregated across the full out-of-sample period.

Because the effective sample decreases with horizon length, the number of available forecasts ranges from 116 (for horizon  $h = 1$ ) to 105 (for  $h = 12$ ) in our dataset of 188 monthly observations. Hyperparameters are tuned by five-fold cross-validation in every rolling window rather than fixed ex ante, allowing the models to adapt to evolving data patterns.

The six-year window balances two considerations: shorter windows yield unstable estimates, while longer windows (e.g. ten years as in Maehashi and Shintani, 2020) reduce the number of available forecasts. Sensitivity checks showed that results were robust across windows of 6–10 years, supporting our choice of 72 months.

**Cross-validation and hyperparameter selection.** Hyperparameters are selected by five-fold cross-validation (CV) within each rolling window. Bergmeir et al. (2018) validate the statistical reliability of this method. The in-sample data are partitioned into five folds of roughly equal size, with four folds used for training and the fifth for validation. This procedure is repeated so that each fold serves as the validation set once. The selection criterion is the mean squared error (MSE).

Five-fold CV is chosen as a compromise between training and validation size. With 72 observations per window, each fold contains about 58 training and 14 validation observations. Using ten folds, for instance, would reduce validation sets to fewer than eight points, producing unstable estimates. We also run the exercise using 115 observations, and our results are robust to the changes in window size. This design therefore provides reliable tuning while preserving sample efficiency.

Given the limited number of observations per training window (72), hyperparameter grids were deliberately kept narrow to avoid overfitting and computational instability. Very large search spaces tend to select overly complex models in small samples, while excessively small grids risk omitting well-performing specifications. The grids we adopt reflect a compromise: broad enough to capture meaningful variation, but restricted to settings consistent with the data constraints.

### 3.2 Forecasting Methods

All computations are conducted in RStudio. As benchmarks, we employ the Autoregressive Integrated Moving Average (ARIMA) and the Random Walk (RW) models following Pratap and Sengupta (2019), Mahajan and Srinivasan (2019), Medeiros et al. (2021) and Chakraborty and Joseph (2017). The Random Walk is defined as:

$$y_{t+h} = y_t + \epsilon_{t+h},$$

where  $y_{t+h}$  represents the forecast of the target variable at time  $t + h$ ,  $y_t$  is the last observed value of the target variable at time  $t$ , and  $\epsilon_{t+h}$  is the error term at time  $t + h$ . ARIMA are estimated in the general form

$$\phi(B)(1 - B^d)y_t = c + \theta(B)\epsilon_t,$$

where  $\epsilon_t$  is a white noise error term,  $d$  is the order of integration,  $B$  represents the lag operator, and  $\phi$  and  $\theta$  are the AR and MA polynomials of order  $p$  and  $q$ . At each iteration of the rolling window, the specification is selected automatically with the “auto.arima()” function of the *forecast* package, using the Akaike Information Criterion (AIC) as the selection criterion. We use the baseline setup of “auto.arima()” as manual specifications did not qualitatively change our results. At each rolling window iteration, the better performing of the ARIMA or RW models serves as the benchmark.

To complement the univariate benchmarks and address concerns that ML gains might reflect comparisons against overly simple models, we also include a baseline Phillips curve estimated on monthly data for Slovakia. Let  $\pi_t$  denote year-over-year HICP inflation (our target) and let  $u_t$  denote the seasonally adjusted registered unemployment rate. The specification is

$$\pi_t = \alpha + \beta u_t + \epsilon_t.$$

The implemented machine learning methods address two challenges in macroeconomic forecasting according to Coulombe et al. (2022). Lasso regression is used for regularization and variable selection in high-dimensional datasets, while Random Forest and Boosting methods capture nonlinear interactions. Finally, a hybrid Lasso–Random Forest approach is proposed to combine these strengths.

**Lasso.** The Lasso of Tibshirani (1996) solves the penalized least squares problem

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \beta)^2 + \lambda \|\beta\|_1 \right\},$$

where  $\lambda$  is the penalty parameter controlling the degree of shrinkage. In each rolling window,  $\lambda$  is selected by five-fold cross-validation from a grid of 100 logarithmically spaced values in  $[10^{-3}, 10^1]$ . This grid is broad enough to accommodate nearly unpenalized models as well as strongly penalized solutions. Strong penalization is crucial in our setting, where each estimation window contains 72 observations but - in the real macro variable setting - more than 200 predictors, as it reduces overfitting risk. Five-fold cross-validation provides a reliable balance between training and validation sample sizes (approximately 58 training and 14 validation observations per fold), avoiding the instability that smaller folds would create. The chosen grid is consistent

with common practice in the forecasting literature (Tibshirani, 1996; Medeiros et al., 2021). The penalty parameter  $\alpha$  is fixed at 1, corresponding to pure Lasso, which yields exact zero coefficients and performs variable selection. Estimation is carried out with the *glmnet* package.

**Random Forest.** Introduced by Breiman (2001) and elaborated on by Coulombe et al. (2021), Random Forest predictions are obtained as the average of  $B$  decision trees

$$\hat{y}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x),$$

where each tree  $T_b$  is grown on a bootstrap sample with a random subset of predictors at each split. We implement 500 trees, which is sufficient to stabilize forecasts in repeated runs without imposing excessive computational cost. Smaller numbers (30–80) gave qualitatively similar results. Random Forest performance depends critically on the number of predictors sampled at each split (*mtry*). In short samples, large *mtry* values tend to produce highly correlated trees and unstable forecasts. Here, *mtry* is tuned in each window by five-fold cross-validation over the range  $1, \dots, \min(60, p)$ , where  $p$  is the number of available predictors. Restricting (*mtry*) to at most 60 reduces computation and avoids overfitting while still allowing broad flexibility. In short samples such as ours, smaller values of (*mtry*) improve generalization by decorrelating the trees, whereas large values tend to produce unstable estimates. The model is implemented using the *randomForest* package in R.

**Boosting.** Introduced by Breiman (2001), Boosting iteratively refines predictions by fitting trees to the residuals from previous steps. With learning rate  $\eta$ , the update at iteration  $m$  is

$$\hat{f}_m(x) = \hat{f}_{m-1}(x) + \eta T_m(x),$$

where  $T_m(x)$  is the tree fitted to the pseudo-residuals at step  $m - 1$ . We implement gradient boosting with the “gbm()” method of the *caret* package. Hyperparameters are selected by five-fold cross-validation within each rolling window. To avoid overfitting in small samples, the grid is restricted to tree depths of  $\{1, 3, 5\}$ , ensuring shallow trees consistent with 72 training observations. Learning rates of  $\{0.01, 0.1\}$  provide gradual updates suitable for small datasets. Minimum terminal node sizes  $\{10, 20\}$  avoid splits based on very few observations, which would not generalize. The number of iterations is fixed at 100, as higher values do not materially improve validation accuracy in our dataset. These restrictions ensure that each tree is shallow and regularized, preventing variance inflation in our dataset. The learning rate controls the pace of improvement: small values  $\{0.01, 0.1\}$  promote gradual refinement, which is better suited to limited data than more aggressive updates.

**Hybrid Lasso – Random Forest.** Finally, we propose a hybrid model inspired by Medeiros et al. (2021). In the first step, Lasso regression is applied to select predictors by shrinking irrelevant coefficients to zero. In the second step, a Random Forest is estimated on the reduced set, with *mtry* tuned as described above. This two-

stage design leverages the strengths of both methods: Lasso addresses the curse of dimensionality by reducing the predictor space, while Random Forest captures nonlinearities and interactions among the retained variables. This design directly addresses the small-sample/high-dimensional challenge: with only 72 observations, fitting a Random Forest to the full predictor set risks severe overfitting, but restricting the feature set via Lasso yields stable results.

### 3.3 Performance Evaluation

Forecast accuracy is assessed using both the Mean Squared Error (MSE) and the Mean Absolute Error (MAE). The MSE is defined as

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2$$

while the MAE is given by

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|.$$

Both measures are reported relative to the benchmark model, which is normalized to one. This relative scaling simplifies interpretation: values below one indicate superior performance relative to ARIMA/RW, while values above one indicate inferior performance. Using both MSE and MAE ensures robustness, as MSE is sensitive to large forecast errors whereas MAE provides a check that results are not driven by outliers.

To evaluate the statistical significance of differences in forecast accuracy, we employ the Diebold–Mariano (DM) test of Diebold and Mariano (2002), written as

$$DM = \frac{\bar{d}}{\hat{\sigma}_d},$$

where  $\bar{d}$  is the average difference between the forecast errors of the compared models and  $\hat{\sigma}_d$  is the standard error of the difference, adjusted for autocorrelation and heteroscedasticity using the Newey-West estimator. This choice is motivated by the time-series nature of our data, where forecast errors are likely serially correlated.

#### 4. Results

In this section, we present and discuss our findings. Table 1 reports the relative Mean Squared Error (MSE)<sup>3</sup> values for all forecasting methods, with the MSE of the benchmark model normalized to one for ease of comparison. The results show that each machine learning technique consistently outperforms the benchmark across all horizons, confirming their superior predictive accuracy. For completeness, we also report the relative MSE of a baseline Phillips curve model. Its performance is notably worse than the benchmark, aligning with prior evidence on the limited forecasting power of Phillips curve specifications, and therefore we do not consider it further in the subsequent analysis.

Initially, the effectiveness of Lasso highlights the significance of applying regularization to the input dataset. Specifically, at the two shortest forecast intervals, Lasso records a relative MSE of approximately 0.55, indicating the most substantial improvement at these intervals. As the forecast horizon extends, the relative MSE slightly rises to about 0.66, and further to 0.74 at  $h = 12$ . Despite this gradual increase, the performance of Lasso markedly exceeds the benchmark across all considered horizons. This robust performance is anticipated, given the presence of multiple correlated variables within the same dataset, which tend to provide overlapping information. Moreover, the shorter time series used in our study compared to other machine learning research might compel Lasso to disregard some variables that are less critical, due to the smaller number of available observations.

Additionally, the significance of nonlinearities and interactions among variables is underscored by the favorable outcomes achieved with nonlinear machine learning methods in forecasting inflation for Slovakia. Both the Boosting and Random Forest methods consistently exceed the performance of the benchmark across all forecast horizons, though the extent of their superiority varies. The relative MSE values for Boosting closely align with those of Lasso, suggesting that even in smaller datasets, such as those typically associated with smaller economies, the absence of regularization does not prevent the methods from identifying critical data relationships. However, the performance of Random Forest is slightly less impressive. It achieves relative MSE values of 0.66 and 0.68 at the shortest forecast horizons. As the horizon lengthens, these values deteriorate, ranging from 0.82 to 0.92. From this, we conclude that boosted decision trees generally offer better predictive accuracy than Random Forest, though both methods still perform better than the benchmark.

---

<sup>3</sup> We divide the MSE values of the machine learning and Phillips curve models by the MSE value of the benchmark. If the resulting value is smaller than 1, it indicates that machine learning models outperform the benchmark.

**Table 1 Relative MSE of Forecasting Methods Based on Macroeconomic Data**

<i>Method</i>	<i>h = 1</i>	<i>h = 3</i>	<i>h = 6</i>	<i>h = 9</i>	<i>h = 12</i>
<i>Benchmark</i>	1	1	1	1	1
<i>Phillips curve</i>	1.859	1.911	2.450	2.492	2.709
<i>Lasso</i>	0.545	0.545	0.663	0.669	0.740
<i>Boosting</i>	0.547	0.530	0.660	0.672	0.746
<i>Random Forest</i>	0.663	0.678	0.822	0.839	0.916

*Source:* authors' calculations

Table 2 presents the relative Mean Absolute Error (MAE) values for each method evaluated. These relative MAE values reinforce the strength of our results, as they are all below one, confirming that the superior performance of machine learning methods is not influenced by occasional large forecast errors. A few notable observations emerge from the analysis. First, the relative MAE values for Lasso and Random Forest are quite similar, indicating comparable sensitivity to forecast errors. Conversely, when considering relative MSE, Boosting and Lasso show akin performance. Furthermore, the extent of outperformance based on relative MAE is less pronounced for the Random Forest and Lasso compared to Boosting. This suggests that the former two are more susceptible to larger forecast errors within our dataset, whereas Boosting manages these errors more effectively.

To assess whether the superior performance of machine learning models is concentrated in specific sub-periods, we also examine the evolution of forecast errors over time (see Appendix Figure B1). The results indicate that both Lasso and Boosting consistently outperform the benchmark throughout the sample. The gains are visible in relatively stable periods as well as during episodes of heightened volatility such as the COVID-19 pandemic, although forecast errors naturally rise for all models in turbulent times. Importantly, no single episode drives the results, suggesting that the advantage of machine learning methods reflects broad improvements in predictive accuracy rather than period-specific effects.

**Table 2 Relative MAE of Forecasting Methods Based on Macroeconomic Data**

<i>Method</i>	<i>h = 1</i>	<i>h = 3</i>	<i>h = 6</i>	<i>h = 9</i>	<i>h = 12</i>
<i>Benchmark</i>	1	1	1	1	1
<i>Lasso</i>	0.766	0.768	0.851	0.853	0.898
<i>Boosting</i>	0.679	0.667	0.766	0.753	0.786
<i>Random Forest</i>	0.763	0.769	0.838	0.837	0.875

*Source:* authors' calculations

The machine learning methods demonstrate robust superiority as evidenced by their outperformance in both Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics. This consistent outperformance across metrics allows us to evaluate the statistical significance of the results using the Diebold-Mariano (DM) test, as detailed in the methodological chapter. The DM test focuses on analyzing the original forecast errors rather than the derived MSE or MAE values. Table 3 displays the outcomes of this test. Notably, numbers marked with two asterisks (\*\*) indicate

horizons where the forecasts of machine learning methods statistically significantly outperform the forecasts of the benchmark model at the 5% significance level.

**Table 3 DM Test p-values**

<i>Method</i>	<i>h = 1</i>	<i>h = 3</i>	<i>h = 6</i>	<i>h = 9</i>	<i>h = 12</i>
<i>Lasso</i>	0.005**	0.005**	0.003**	0.005**	0.019**
<i>Boosting</i>	0.034**	0.040**	0.101	0.130	0.296
<i>Random Forest</i>	0.016**	0.016**	0.040**	0.046**	0.122

Source: authors' calculations

The outperformance of Lasso is statistically significant at four horizons, which suggests that by fine-tuning hyperparameters at each step, this method effectively selects the most relevant variables for the forecasts to surpass the benchmark model. Boosting, the best-performing method overall, shows statistically significant superiority across all horizons, underscoring the importance of nonlinearities in the data generation process. This finding is particularly relevant as there are no other published studies addressing potential nonlinearities in Slovakia's data in their forecasting models, as highlighted in our introduction. Our results advocate for the inclusion of nonlinear data-driven models in forecasting, given their ability to capture significant nonlinear effects. Conversely, Random Forest achieves significant outperformance only on two horizons. Despite the increasing accuracy with more trees, Boosting remains more computationally efficient and easier to optimize via cross-validation compared to Random Forest. Given the steep rise in computational demand with additional trees in the Random Forest, we adopt a hybrid approach as detailed in the methodological section to maintain efficiency. The outcomes of this approach are documented in Table 4.

**Table 4 % Improvement of Relative MSE Compared to the Benchmark**

<i>Method</i>	<i>h = 1</i>	<i>h = 3</i>	<i>h = 6</i>	<i>h = 9</i>	<i>h = 12</i>
<i>Lasso</i>	45.45%	45.43%	33.70%	33.05%	25.91%
<i>Boosting</i>	45.29%	46.98%	33.96%	32.74%	25.40%
<i>Random Forest</i>	33.61%	32.12%	17.77%	16.05%	8.39%
<i>Lasso + Random Forest</i>	46.31%	48.94%	36.62%	34.77%	30.23%

Source: authors' calculations

To clearly demonstrate the effectiveness of the models, we not only present the relative Mean Squared Error (MSE) and Mean Absolute Error (MAE) values but also illustrate the extent of outperformance in percentage terms in Table 4. As anticipated, both Lasso and Boosting significantly outperform the benchmark, exceeding it by more than 45% on several horizons. However, the Random Forest shows less impressive outperformance in percentage terms, which is also statistically insignificant. Table 4 additionally details the outcome of using a hybrid approach that combines Lasso and Random Forest. Initially, we employ Lasso to regularize the dataset in each iteration, effectively reducing unimportant coefficients to zero. Subsequently, we apply Random Forest to the variables selected by Lasso. This method not only markedly enhances performance over the traditional Random Forest but also emerges as the top-performing method in terms of percentage improvements. This efficiency likely stems

from the reduction in the number of variables, which allows for achieving comprehensive coverage of critical nonlinearities and variable interactions with fewer trees. Furthermore, as noted by Medeiros et al. (2021), when the Random Forest model is applied to a set of regularized variables, its capacity to discern important variable interactions is notably amplified, but it becomes harder to capture nonlinear relationships. In such case, if the goal is to explore the effects of nonlinearities further, one might consider other nonlinear methods like Boosting, which are specifically designed to handle such complexities effectively.

In conclusion, our study demonstrates that machine learning methods can significantly enhance the accuracy of inflation forecasts for Slovakia over the benchmark model. We achieve this improvement by either optimizing the number of input variables through regularization techniques such as Lasso, or by utilizing models that capture nonlinear dynamics, such as Boosting. The forecasting performances of these methods are quite comparable, suggesting that nonlinear relationships do play a crucial and statistically significant role. Moreover, regularization not only streamlines computation but also boosts forecasting accuracy. For the best results, combining both regularization and nonlinear predictive methods is advisable. This hybrid approach yields the most precise inflation forecasts in our tests. However, it is important to acknowledge that machine learning methods at present largely remain "black boxes", as highlighted by Masini et al. (2021). Masini et al. (2021) also state that although various interpretative techniques are available, there is no consensus within the academic community on their adequacy, especially when compared to more traditional models like VAR, which allow for straightforward interpretation through impulse responses or variance decompositions.

**Table 5 Relative MSE of Forecasting Methods Based on Soft Indicators**

<i>Method</i>	<i>h = 1</i>	<i>h = 3</i>	<i>h = 6</i>	<i>h = 9</i>	<i>h = 12</i>
<i>Benchmark</i>	1	1	1	1	1
<i>Lasso</i>	1.760	1.757	2.158	2.085	2.117
<i>Boosting</i>	1.721	1.719	2.141	1.998	1.981
<i>Random Forest</i>	1.730	1.669	2.087	1.969	1.999

Source: authors' calculations

To conclude our analysis, we evaluate the performance of machine learning methods on our second dataset, which consists of soft indicators. Coulombe et al. (2022) suggest utilizing soft indicators as inputs for machine learning models. We assess these methods by examining the relative Mean Squared Error (MSE) in Table 5.

In summary, soft indicators are found to be ineffective for forecasting inflation in Slovakia. Table 5 indicates that all relative Mean Squared Error (MSE) values are greater than one, signifying that a simple univariate benchmark model produces more accurate forecasts than the complex machine learning methods utilizing soft indicators.

Given the weak predictive performance of soft indicators in isolation, we expect that their inclusion alongside hard macroeconomic data will either leave the forecasting accuracy of machine learning models unchanged or even worsen it. These expectations are confirmed by the results reported in Table B1. Lasso, which performs variable selection, remains essentially unaffected: given the poor performance of soft

indicators, it does not select any of them and thus stays robust. Boosting exhibits only very small changes in relative MSE (for instance, moving from 0.660 to 0.655 or from 0.746 to 0.831). By contrast, the performance of the Random Forest model deteriorates slightly at every horizon. Overall, the inclusion of soft indicators introduces additional complexity and substantially increases computing time, while either adding no value or marginally worsening the performance of the models.

Despite the greater accessibility of soft indicators, their use alone is not recommended for reliable inflation forecasting in Slovakia. Neither it is recommended to combine them with hard macroeconomic data. Instead, selected machine learning methods applied to observed macroeconomic data for inflation forecasting have demonstrated considerable superiority in our sample, substantially outperforming the benchmark model. These benefits make machine learning techniques with observed macroeconomic data the preferred choice for such economic forecasting tasks.

## 5. Conclusions

In this study, we look at how well machine learning techniques can predict inflation rates in Slovakia, especially during the unstable economic times caused by the COVID-19 pandemic. In our sample, we find that machine learning methods like Lasso, Boosting, and Random Forest are much better at forecasting inflation than traditional models.

Lasso is useful because it simplifies the models by selecting only the most important variables. Boosting and Random Forest tackles the complex, nonlinear patterns in the data, giving more accurate and reliable forecasts. These methods show their strength by consistently providing lower error rates in predictions compared to standard approaches.

Our findings are important for policymakers, but their interpretation must be made with care. While machine learning methods deliver superior predictive accuracy, they largely remain “black boxes” and do not offer the same causal explanations as traditional structural econometric models, such as VARs with impulse response analysis. As such, they should be seen as a complement to rather than a replacement for models that provide clear narratives about the transmission mechanisms of shocks.

On the other hand, the availability of additional information can significantly reduce uncertainty in decision-making processes. By incorporating this supplementary data, the accuracy of forecasts improves, leading to more reliable predictions. Consequently, better forecasts enable more informed and effective decisions (both short- and medium-term), ultimately enhancing the overall quality of the decision-making process.

In summary, this paper shows that in our new sample, machine learning can offer more precise and reliable ways of predicting Slovakian economic trends like inflation than traditional methods. When used alongside structural models, these methods can enrich the information set available to policymakers and improve their situational awareness. Further studies could expand these methods to other economic indicators, explore interpretability techniques, and verify whether these gains hold across other small open economies.

## REFERENCES

- Athey S (2019): The Impact of Machine Learning on Economics. In Agrawal A, Gans J, Goldfarb A (eds): *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, Chicago, 507–547.
- Atkeson A, Ohanian L (2001): Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve bank of Minneapolis Quarterly Review*, 25:2–11.
- Bergmeir C, Hyndman RJ, Koo B (2018): A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction. *Computational Statistics and Data Analysis*, 120:70–83.
- Bolhuis AM, Rayner B (2020): Deus ex Machina? A Framework for Macro Forecasting with Machine Learning. *IMF Working Paper WP/20/45*.
- Breiman L (2001): Random Forests. *Machine Learning*, 45(1):5–32.
- Chakraborty C, Joseph A (2017): Machine Learning at Central Banks. *Bank of England Working Paper No. 674*.
- Chen Y-C, Turnovsky S, Zivot E (2014): Forecasting Inflation Using Commodity Price Aggregates. *Journal of Econometrics*, 183(1):117–134.
- Coulombe PG, Marcellino M, Stevanović D (2021): Can Machine Learning Catch the Covid-19 Recession? *National Institute Economic Review*, 256(1): 71–109.
- Coulombe PG, Leroux M, Stevanović D (2022): How is Machine Learning Useful for Macroeconomic Forecasting? *Journal of Applied Econometrics*, 37(5):920–964.
- Diebold FX, Mariano RS (2002): Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 20(1):134–144.
- Forni M, Hallin M, Lippi M, Reichlin L (2003): Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area? *Journal of Monetary Economics*, 50(6):1243–1255.
- Giannone D, Lenza M, Primiceri GE (2021): Economic Predictions with Big Data: The Illusion of Sparsity. *European Central Bank Working Paper No 2542*.
- Groen J, Paap R, Ravazzolo F (2013): Real-Time Inflation Forecasting in a Changing World. *Journal of Business and Economic Statistics*, 31(1):29–44.
- Jung J, Patnam M, Ter-Martirosyan A (2018): An Algorithmic Crystall Ball: Forecasts Based on Machine Learning. *IMF Working Paper 2018/230*.
- Maehashi, K., Shintani, M. (2020). Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan. *Journal of The Japanese and International Economies*, 58:1–17.
- Mahajan K, Srinivasan A (2019): Inflation Forecasting in Emerging Markets: A Machine Learning Approach. *CAFRAL Working Paper*.
- Masini RP, Medeiros MC, Mendes EF (2021): Machine Learning Advances for Time Series Forecasting. *The Journal of Economic Surveys*, 37(1):76–111.
- Medeiros MC, Mendes E (2016): Regularization of High-Dimensional Time-Series Models with Non-Gaussian and Heteroskedastic Errors. *Journal of Econometrics*, 191(1):255–271.
- Medeiros MC, Vasconcelos G, Veiga A, Zilberman E (2021): Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business and Economic Statistics*, 39(1):1–45.
- Mullainathan S, Spiess J (2017): Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):574–589.
- Nechvátalová L (2024): Multi-Horizon Equity Returns Predictability via Machine Learning. *Czech Journal of Economics and Finance*, 74(2):142–190.
- Pratap B, Sengupta S (2019): Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts?. *RBI Working Paper Series. No. 04/2019*.

- Tibshirani R (1996): Regression Shrinkage and Selection Via Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288.
- Vargas AR (2020): Forecasting Costa Rican Inflation with Machine Learning Methods. *Latin American Journal of Central Banking*, 1(4):1–21.
- Zhang GP, Qi M (2005): Neural Network Forecasting for Seasonal and Trend Time Series. *European Journal of Operational Research*, 160(2):501–514.