Comparing Machine Learning Models for US Inflation Forecasting: Evidence from FRED-MD

Silvia Ianeselli Emilia Infante Ece Taşan Özel
Orhun Özel

Barcelona School of Economics

December 8, 2025

Abstract

This study examines the performance of machine learning methods for forecasting US inflation in a data-rich environment. Using the FRED-MD database with 127 macroeconomic and financial variables, we compare the forecasting accuracy of LASSO, Ridge regression, Elastic Net, and Random Forest (RF) models against traditional benchmarks including the Rolling Sample Mean, Random Walk (RW) and an autoregressive AR(4) model. Following the methodology of Medeiros et al. (2021), we employ a rolling window framework to generate out-of-sample forecasts for 1-month and 3-month ahead horizons across two periods: 2001–2015 and 2015–2025. Our results indicate that machine learning methods consistently outperform traditional benchmarks, with Random Forest achieving the lowest root mean squared error (RMSE) for 3-month ahead forecasts, representing improvements of up to 20–30% over the AR(4) benchmark. For 1-month ahead forecasts, regularization methods, particularly Elastic Net, demonstrate superior performance. Variable importance analysis reveals that while Ridge and AR models rely heavily on autoregressive terms and price indices, Random Forest exploits nonlinear relationships among employment, housing, and interest rate variables. These findings suggest that the combination of rich datasets with machine learning methods offers substantial gains for inflation forecasting, and that both regularization techniques and nonlinear specifications play crucial roles in improving forecast accuracy.

Keywords: Inflation forecasting, Machine learning, LASSO, Ridge regression, Elastic Net, Random Forest, FRED-MD

1 Introduction

Inflation forecasting stands as one of the most crucial yet challenging tasks in macroeconomic analysis. Forecasting consumer price inflation accurately in the near and medium term has considerable implications for monetary policy, other policy choices, and business decisions in the wider economy (Joseph et al., 2024). Accurate inflation forecasts are crucial for central banks to design appropriate and timely policy responses and to communicate the path at which inflation is expected to return to target.

Inflation forecasting is of great value to households, businesses and policymakers (Medeiros et al., 2021). In emerging markets particularly, producing reliable inflation forecasts is a constant challenge for policymakers and of greatest importance to economic agents and their investment decisions. Inflation adds uncertainty to investment decisions and shortens the investment horizon, especially in emerging markets, making the construction of accurate forecasts a relevant issue in such economies (Araujo and Gaglianone, 2023). The mandate of the European Central Bank (ECB) is to maintain price stability over the medium-term, and inflation projections play a pivotal role in shaping monetary policy decisions (Lenza et al., 2025).

Despite the great benefits of forecasting inflation accurately, improving simple benchmark models has been proven to be a major challenge for both academics and practitioners (Medeiros et al., 2021). Building accurate forecasts is generally not an easy task because it requires an approach complex enough to incorporate relevant variables but also focused on excluding irrelevant data (Araujo and Gaglianone, 2023).

Literature shows that it is hard to beat simple autoregressive benchmarks significantly, even when exploiting large data, and uncertainty in the forecasts is relatively wide, with considerable variation across models and horizons (Joseph et al., 2024). Modeling inflation dynamics remains a significant challenge, given the multitude of potential driving factors, the complexity of their interactions with inflation, and the relatively short historical sample available for econometric analysis (Lenza et al., 2025).

Machine learning (ML) methods, in general, are able to identify nonlinear patterns in the data, hidden to standard linear models, thus offering a compelling alternative approach to traditional econometric models (Araujo and Gaglianone, 2023). The application of machine learning to inflation forecasting has gained substantial momentum. Despite previous skepticism, recent evidence shows that it is possible to beat the usual univariate benchmarks for inflation forecasting (Medeiros et al., 2021). The expanding literature highlights the importance of accounting for non-linearities in the dynamics of key policy-relevant variables, particularly for the robustness of risk assessments conducted by central banks (Lenza et al., 2025). A central debate in the literature concerns whether the relationship between inflation and its determinants is best characterized as linear or non-linear, with important implications for both forecasting accuracy and monetary policy assessment (Lenza et al., 2025).

This study focuses on four classes of machine learning methods that represent different approaches to the forecasting problem. Following Medeiros et al. (2021), we use the FRED-MD database, which contains 127 monthly macroeconomic and financial variables covering nine economic categories, to forecast US CPI inflation. The models used are LASSO (Least Absolute Shrinkage and Selection Operator), Ridge regression, Elastic Net, and Random Forest. Each method offers distinct advantages and represents different approaches about how to extract predictive information from large datasets. We compare the forecasting performance of these four ML methods against an Ordinary Least Squares (OLS) benchmark.

2 Data Description

We employ the FRED-MD database, a macroeconomic dataset developed by McCracken and Ng (2016). The dataset provides a monthly panel of U.S. macroeconomic and financial indicators, grouped into major economic categories such as output and income, labor market conditions, housing, consumption, orders and inventories, money and credit, interest rates, prices, and stock markets.

Following the transformation procedures outlined in Medeiros et al. (2021), each series is converted into a stationary representation using variable-specific transformation codes. After applying these transformations, we restrict the sample to the 1960–2024 period and address missing observations (the transformation and missing observation strategies are described in detail in the next section). The final dataset consists of a balanced monthly panel with 780 observations and 127 transformed macroeconomic indicators, where inflation (log-difference of CPI) serves as the target variable in the forecasting exercise.

3 Feature Engineering and Descriptive Statistics

3.1 Data Transformation

Following Medeiros et al. (2021), all variables in the FRED-MD database undergo transformations to ensure stationarity, a critical requirement for time series forecasting models. The transformation applied to each variable is determined by its statistical properties and is denoted by a transformation code (*tcode*).

The transformation codes range from 1 to 7, designed to induce stationarity. A *tcode* of 1 indicates that no transformation is required, as the variable is already stationary in levels. This applies primarily to variables such as interest rates and survey-based indices. A *tcode* of 2 applies the first difference operator, computing the change between consecutive observations. This transformation removes linear trends and is appropriate for variables that exhibit non-stationary behavior in levels but become stationary after differencing. For *tcode* 3 it applies the second difference operator, which computes the difference of first differences. Variables measured in levels that exhibit exponential growth or multiplicative trends are transformed using logarithms (*tcode* 4), which stabilizes variance and converts multiplicative relationships into additive ones.

The most commonly applied transformation is *tcode* 5, which computes the first difference of the natural logarithm of the variable. This transformation yields the growth rate or percentage change and is applied to the majority of macroeconomic series. The inflation measure used as the dependent variable is constructed using this transformation, specifically:

$$\pi_t = \log(P_t) - \log(P_{t-1}) \tag{1}$$

where P_t represents the Consumer Price Index in period t. Similarly, all the price variables are log-differenced one time. *Tcode* 6 applies the second difference of logarithms, useful for variables with persistent growth rates, while *tcode* 7 computes the first difference of the percentage change, capturing acceleration or deceleration in growth rates. The comprehensive set of transformations ensures that all predictor variables exhibit stationary properties suitable for regression-based forecasting methods.

Following Medeiros et al. (2021), we incorporated the first four principal components extracted from the initial set of 126 explanatory variables. We also included up to four lags of each series when estimating all forecast horizons. Finally, to make our results comparable to those reported by Medeiros et al. (2021) for the 2001–2015 period, we introduced a dummy variable capturing the 2008–11 crisis episode.

3.2 Dealing with Missing Values

To handle missing values in the dataset, we employ a Principal Component Analysis (PCA) based imputation method following the approach outlined by McCracken and Ng (2016) for FRED-MD data. This iterative Expectation-Maximization (EM) algorithm begins by standardizing all numeric variables using means and standard deviations calculated from observed values only. Missing values are initially set to zero (the mean of standardized data), and the algorithm then iteratively performs PCA to extract the first five principal components from the complete dataset. In each iteration, the missing values are replaced with their projections onto this lower-dimensional factor space, leveraging correlations across all 126+ variables to inform the imputation. The process continues until convergence (relative change $< 10^{-6}$) or a maximum of 100 iterations is reached. By utilizing the full information set, this method captures common patterns across the entire macroeconomic dataset to produce more accurate and theoretically grounded imputations that preserve the covariance structure of the original data.

3.3 Descriptive Statistics

Figure 1 shows that monthly U.S. inflation moves around a stable average but becomes more volatile in certain periods. Volatility increases clearly in the mid-1970s, during the 2008 financial crisis, and again after 2020.

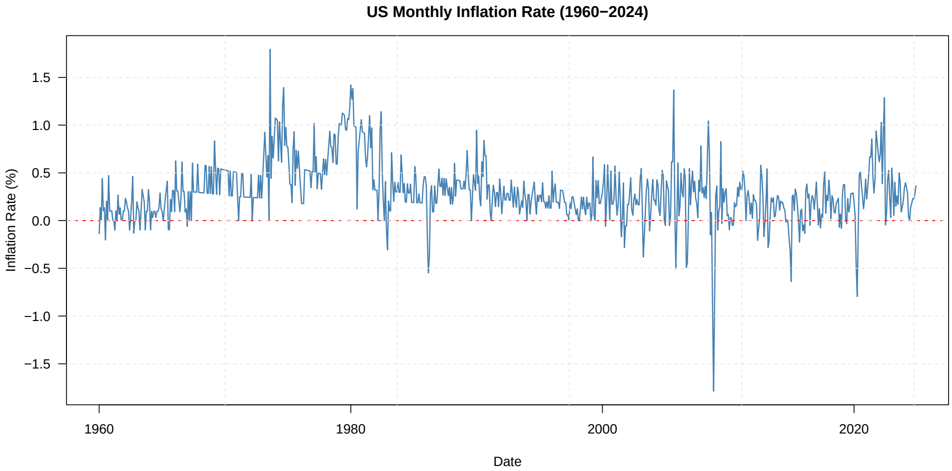


Figure 1: Temporal evolution of US monthly inflation rate

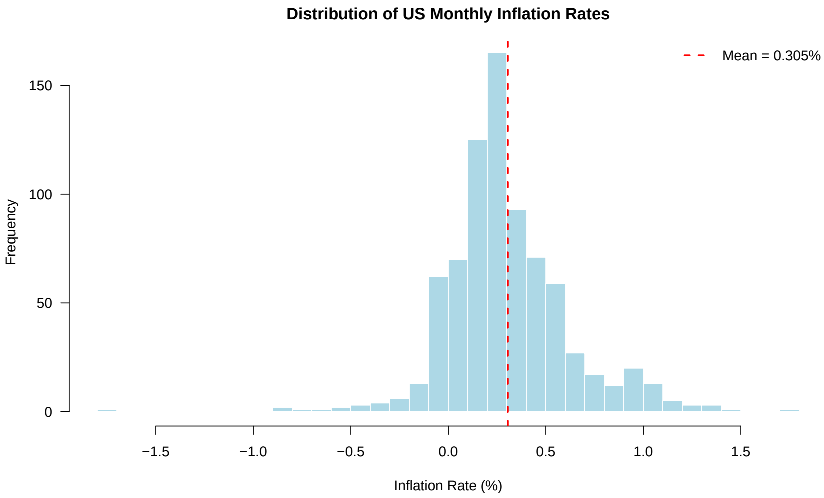


Figure 2: Distribution of monthly inflation rates

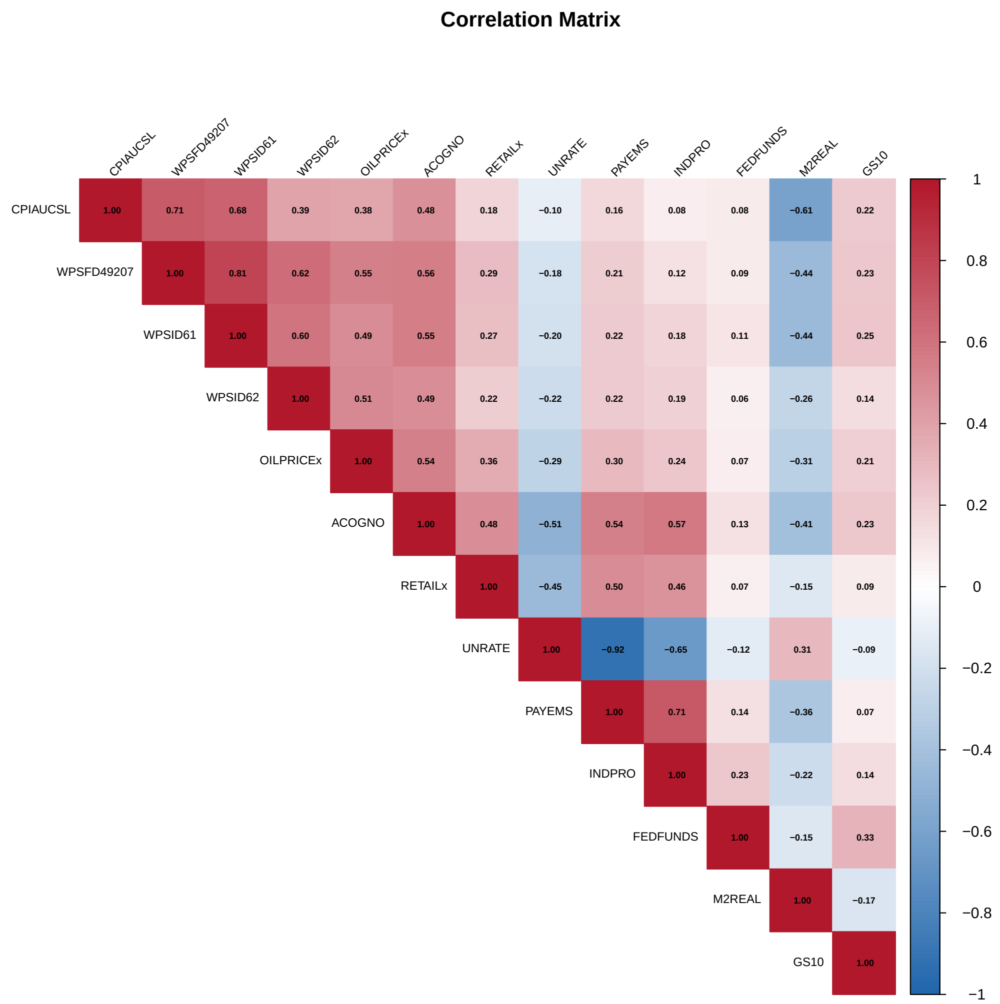
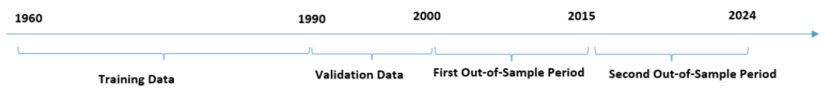


Figure 3: Correlation structure among inflation (CPIAUCSL) and selected macroeconomic predictors

Note: The selected variables: (1) Supply-side factors: producer price indices at different production stages (WPSFD49207, WPSID61, WPSID62) and oil prices (OILPRICEx); (2) Demand indicators: consumer goods orders (ACOGNO) and retail sales (RETAILx); (3) Labor market: unemployment rate (UNRATE) and employment (PAYEMS); (4) Monetary conditions: Federal Funds rate (FEDFUNDS), real money supply (M2REAL), and long-term interest rates (GS10). Industrial production (INDPRO) captures real economic activity.

4 Methodology and Results

The evaluation employs a rolling window out-of-sample forecasting framework. This method is preferred to prevent any data leakage. In other words, this approach ensures that forecasts are generated using only information that would have been available at each point in time. The full sample spans from January 1960 to December 2024, providing 780 monthly observations. The initial 372 months, from January 1960 to December 1990, serve as the training window. The consequent 120 months, from January 1991 to December 2000, serve as the validation window. Model parameters resulting in the minimum RMSE in the validation window are selected to be used in the test data prediction. We divide the out-of-sample period (test data) into two: (i) first period spans years between 2001-2005 (this is the second sample in Madeiros et al.), (ii) and the second period spans years between 2016-2024.This division enables the assessment of whether ML methods demonstrate robust performance across distinct economic regimes.



The rolling window mechanism operates as follows. For each forecast date t in the out-of-sample period, a model is estimated using a fixed-length window of the most recent R_h observations, where R_h denotes the window size that varies according to the forecasting horizon h. After generating the h-step-ahead forecast, the window advances by one period: the oldest observation is dropped and the most recent observation is added. This process continues iteratively until forecasts have been generated for all dates in the out-of-sample period. This study examines thirteen distinct forecast horizons: direct forecasts for h = 1 and 3 months ahead inflation rates. For each horizon, a separate rolling window exercise is conducted, generating a complete sequence of out-of-sample forecasts. Each model is re-estimated at every forecast origin using only the data available up to that point, ensuring that parameter estimates adapt to evolving economic conditions throughout the out-of-sample period. Forecast accuracy is evaluated using the root mean squared error (RMSE), a standard loss function computed over all out-of-sample forecasts. RMSE is calculated as the square root of the average squared forecast error across all T - T + 1 out-of-sample predictions. This metric penalizes large forecast errors more heavily due to the squaring operation and represents the most commonly used measure of forecast accuracy in the literature.

4.1 OLS Benchmark

4.1.1 Random Walk

The first benchmark is the Random Walk (RW) model, in which inflation is forecast simply by carrying forward its most recent observed value. For the first out-of-sample period, the RW model yields an RMSE of 0.335 for one-month-ahead forecasts, increasing to 0.471 when the horizon is extended to three months. In the second out-of-sample period, the RMSE falls to 0.250 for one-month-ahead forecasts and rises to 0.328 for the three-month horizon. As expected, forecast errors increase when the prediction horizon lengthens.

4.1.2 Rolling Sample Mean (RSM)

The Rolling Sample Mean serves as our simplest benchmark model. For each forecast, this method computes the arithmetic mean of all observed inflation values in the training window up to time t. Formally, the forecast for period $t+h$ is given by:

$$\hat{\pi}_{t+h} = \frac{1}{T}\sum_{i=1}^{T}\pi_i \tag{2}$$

where T is the number of observations in the training sample up to time t. This naive benchmark provides a lower bound for forecast performance and represents the assumption that future inflation will equal the historical average. Despite its simplicity, the Rolling Mean is a commonly used benchmark in the inflation forecasting literature, as it is notoriously difficult to beat with more sophisticated models, particularly at shorter horizons.

4.1.3 Autoregressive Model (AR)

We employ an autoregressive model of order 4, denoted AR(4), as our second benchmark. This model assumes that current inflation can be predicted using its own past values. The AR(4) specification is expressed as:

$$\pi_{t+h} = \beta_0 + \beta_1\pi_{t+h-1} + \beta_2\pi_{t+h-2} + \beta_3\pi_{t+h-3} + \beta_4\pi_{t+h-4} + \varepsilon_{t+h} \tag{3}$$

where π_t represents the inflation rate at time t, $\beta_0, \ldots, \beta_4$ are the model parameters estimated via ordinary least squares (OLS), and ε_{t+h} is the error term. The choice of four lags captures the quarterly dynamics often present in monthly macroeconomic data while avoiding over-parameterization.

4.2 Linear Methods

4.2.1 Ridge Regression

Selection of lambda is based on one step ahead out of sample forecast performance in a rolling window setting using training data similar to the approach in LASSO. According to Figure X, the best lambda came out to be 1.57. When it comes to variable importance, some of the most important variables are (1) personal expenditures towards services lag3, (2) personal expenditures towards services lag4, and (3) general personal expenditures lag1. The best variables came from the past values of price movements as expected. The coefficients of the selected variables are far less than that of lasso. Hence while lasso is good at dropping variables, the ridge regression better pulled them towards 0.

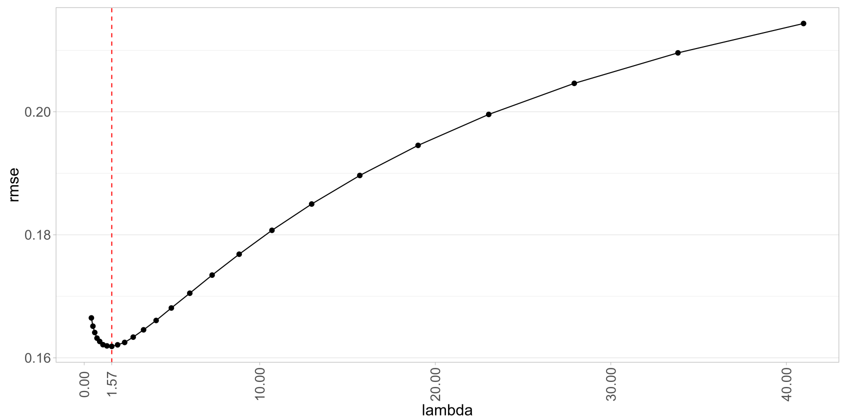


Figure 4: 1-Step Ahead Out of Sample RMSE Performances for Different Values of

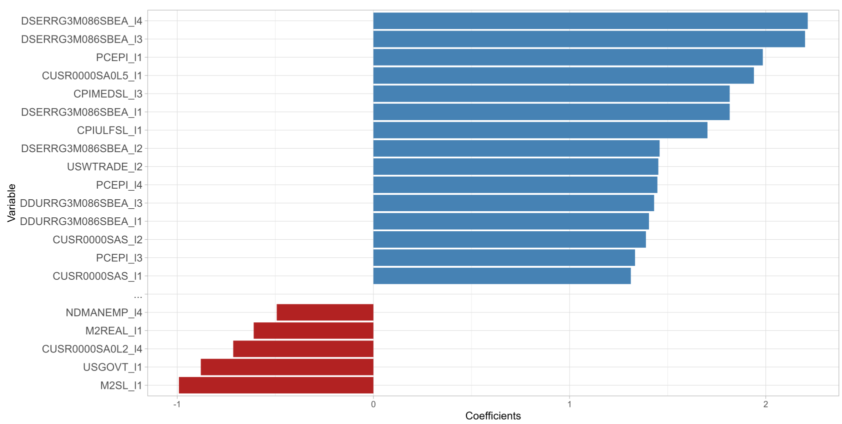


Figure 5: Ridge variable importance

4.2.2 LASSO

Lasso parameter tuning sample is between 1960-2001. The last 10 years of this sample between 1991-2001 is used as a validation sample. Best lambda values are obtained using 1-step ahead out of sample forecast performance in a rolling window setting. According to Figure X, the best lambda came out to be 0.0169. This lambda value picks 38 covariates and the remaining variables' coefficients are set to 0 (Figure X). When it comes to the variable importance, some of the most important variables are in order (1) CPI excluding medical care lag1, (2) personal expenditures towards services lag3 and (3) personal expenditures towards services lag4. The best variables came from the past values of price movements as expected.

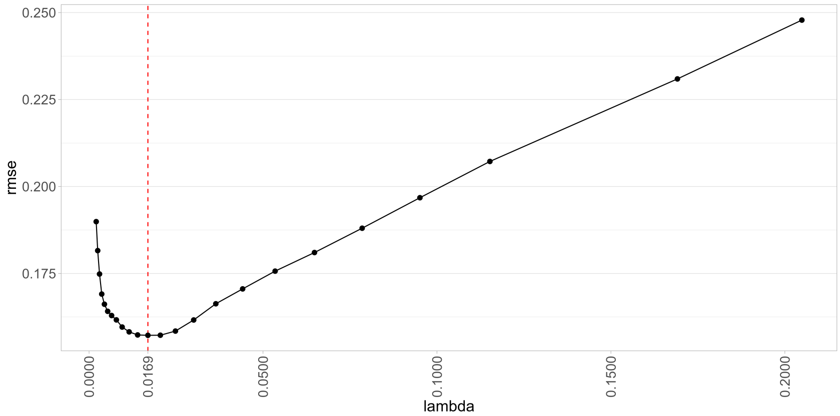


Figure 6: 1-step ahead out of sample RMSE performances for different values of lambda

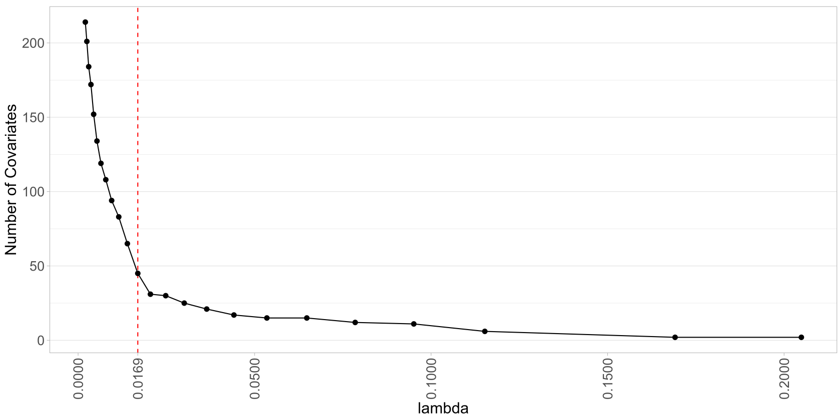


Figure 7: Number of selected covariates for different values of lambda

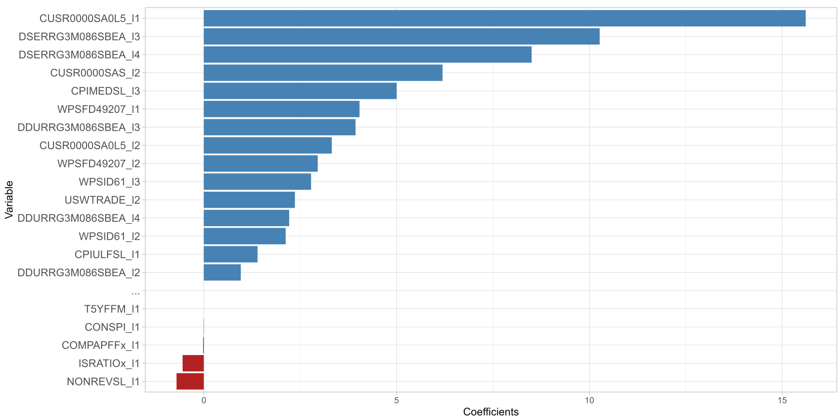


Figure 8: LASSO variable importance

4.2.3 Elastic Net

Selection of lambda and alpha parameters of Elastic Net is also based on one step ahead out of sample forecast performance in a rolling window setting. Here, for every alpha between 0.1–0.2–...0.9, we calculated the best possible lambda. Then using the best lambda for each alpha, we compare their results to find that the best values of alpha and lambda are in order 0.7 and 0.0218 (Table 1). The alpha value is much closer to LASSO parameter of 1, as a result of this the predictions of Elastic Net are comparatively more similar to LASSO predictions. The variable importance ranking also closely resembles the LASSO's chart. However, an important difference can be seen in the magnitude of the coefficients. Here, the coefficient is significantly lower than that of LASSO due to the lower alpha. Compared to pure LASSO, the Elastic Net inherited some part of the inner-workings of the Ridge. The selected number of variables at the end is 47.

Table 1: Elastic Net Tuning Results

Alpha	**Lambda**	**RMSE**
0.1	0.1099	0.160
0.2	0.0764	0.160
0.3	0.0509	0.159
0.4	0.0382	0.159
0.5	0.0306	0.158
0.6	0.0255	0.158
0.7	**0.0218**	**0.158**
0.8	0.0191	0.158
0.9	0.0170	0.159

Note: The optimal combination of alpha = 0.7 and lambda = 0.0218 (highlighted in bold) minimizes the out-of-sample RMSE. The alpha value closer to 1 indicates the model behavior is more similar to LASSO than Ridge regression.

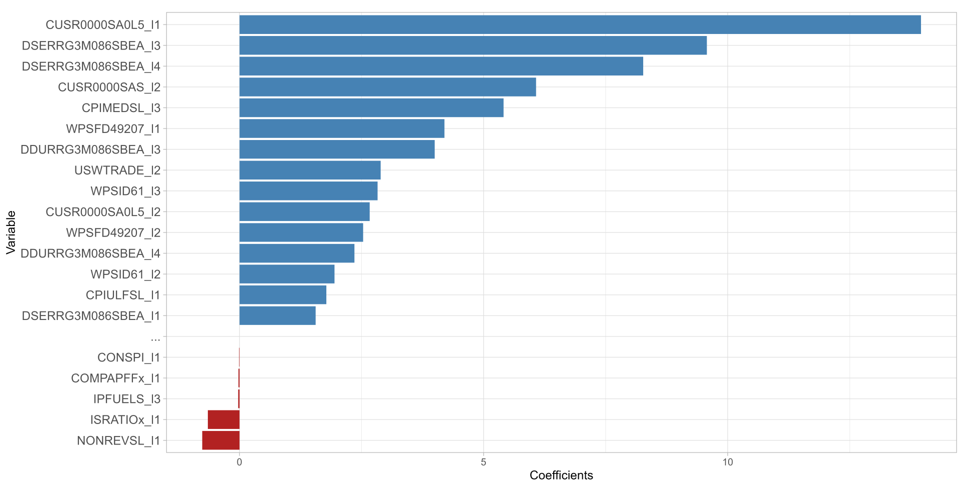


Figure 9: Elastic net variable importance

4.3 Non-linear Analysis

4.3.1 Random Forest

We employ a random forest algorithm which is a non-linear method, in order to forecast US inflation. Before measuring the forecast performance in two out-of-sample periods, we tune the number of variables randomly sampled as candidates at each split (*mtry*), unlike Medeiros et al. (2021) which uses default *mtry* in R. The random forest analysis package in R sets the default *mtry* as one third of the total number of features ($p/3$). In our analysis, we tune *mtry* by doing a grid search using the validation window (1991–2000). We chose 13 candidates for *mtry*: $\{2, 3, 5, 8, 10, 15, 25, p/10, p/8, p/6, p/4, p/3, p/2\}$. These candidates consist of some portions of the total number of features (p) and some very small numbers in order to see if allowing for more variation compensates for the increased bias. The 12 candidates for *mtry* are $\{2, 3, 5, 8, 10, 15, 25, 52, 65, 87, 130, 173, 260\}$, where the total number of features (with four PCA factors and 4 lags of all features) is 520. We achieved the minimum RMSE at $mtry = 52$ in the validation data (Figure 10). As this tuning is made in the validation data which is right before the first out-of-sample data, this *mtry* is actually optimized for the first out-of-sample period (2001–2015). Therefore, we used this value for RF forecasts in the first out-of-sample period. For the second out-of-sample data (2016–2024), it would be ideal to retune the *mtry* variable in a validation data chosen just before the second out of sample data (for example validation data for the second forecast sample could be the period between 2005–2015). However, we used the same *mtry* (52) for both forecast samples, for the sake of simplicity.

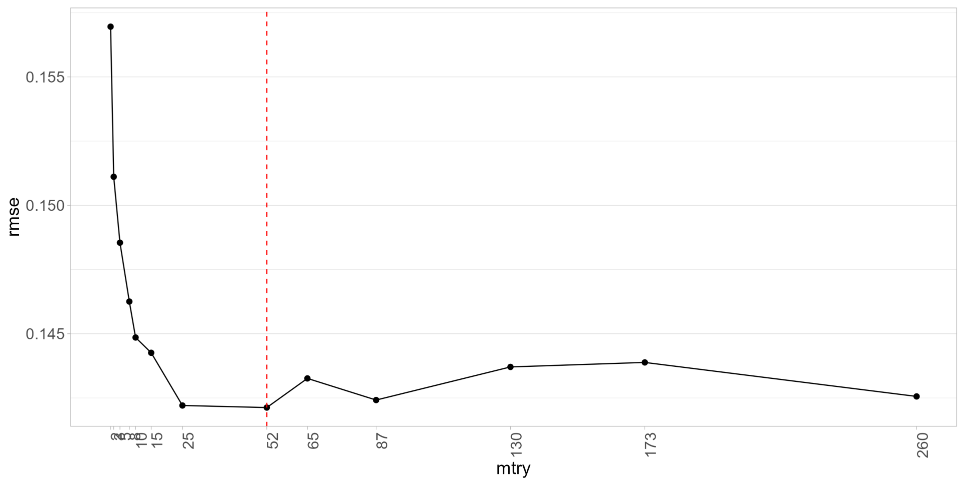


Figure 10: RMSE for different number of variables randomly sampled at RF splits (*mtry*)

After selecting the optimal value of *mtry*, we produced monthly inflation forecasts for two out-of-sample periods. For the first period (2001–2015), the one-month-ahead forecast yielded an RMSE of 0.298. As expected, increasing the forecast horizon to three months led to a higher error, with the RMSE rising to 0.327.

For the second out-of-sample period (2016–2024), forecast accuracy improved. The one-month-ahead RMSE decreased to 0.224, and, as in the earlier period, extending the horizon to three months resulted in a higher RMSE of 0.259. A detailed comparison with alternative forecasting methods is presented in the following section.

4.4 Comparison of Methods

Table 2 presents the forecast performance across all models for both out-of-sample periods and horizons. Machine learning methods consistently outperform traditional benchmarks, with Random Forest achieving the lowest RMSE for 3-step ahead forecasts (0.327 and 0.258 for the two periods), representing approximately 20–30% improvements over the AR(4) benchmark. For 1-step ahead forecasts, regularization methods (LASSO, Ridge, Elastic Net) demonstrate superior performance, with Elastic Net and Random Forest both achieving RMSE of 0.224 in the recent period.

All models show improved accuracy in the second period (2015–2025) compared to the first (2001–2015), likely reflecting the more stable inflation environment outside the 2008 financial crisis period. Random Forest exhibits remarkable stability across horizons, while Ridge regression shows more pronounced performance degradation at longer horizons.

Table 3 validates our methodology against Medeiros et al. (2021) for the 2001–2015 period. Our replication shows strong consistency, with identical RMSEs for LASSO and Elastic Net at the 1-step horizon (0.28) and exact matches for Random Forest at 3-step ahead (0.33). Minor discrepancies in AR and Ridge results likely stem from data revisions in FRED-MD or implementation details. The close replication provides confidence in our extended analysis for the 2015–2025 period.

Table 2: RMSE Comparisons of Different Models

	2001-2015		**2015-2025**	
	1 step	**3 step**	**1 step**	**3 step**
RW	0.335	0.471	0.250	0.328
RSM	0.359	0.359	0.283	0.283
AR	0.369	0.356	0.279	**0.257**
Ridge	**0.299**	0.375	**0.226**	0.282
Lasso	0.279	0.341	0.226	0.263
ElNet	0.278	0.348	0.224	0.269
RF	**0.298**	**0.327**	**0.224**	**0.258**

Note: Root Mean Squared Error (RMSE) for different forecasting models across two out-of-sample periods and two forecast horizons. Lower values indicate better forecast performance. Best performing models for each period-horizon combination are shown in bold. RW = Random Walk, RSM = Rolling Sample Mean, AR = Autoregressive model.

Table 3: RMSE Comparisons for 2001-2015 Between Author's Calculations and Medeiros (2021)

	1 Step Ahead		**3 Step Ahead**	
	Current Results	**Medeiros (2021)**	**Current Results**	**Medeiros (2021)**
RW	0.34	0.34	0.47	0.47
AR	0.37	0.31	0.36	0.37
Ridge	**0.30**	**0.29**	0.36	**0.34**
Lasso	**0.28**	**0.28**	0.34	0.35
ElNet	**0.28**	**0.28**	0.35	0.34
RF	**0.30**	**0.29**	**0.33**	**0.33**

Note: Comparison of RMSE values between our replication and the original Medeiros et al. (2021) results for the 2001-2015 period. Values in bold indicate close replication (difference ≤ 0.02). Our results are generally consistent with the original study, with some minor discrepancies potentially due to data revisions or implementation details.

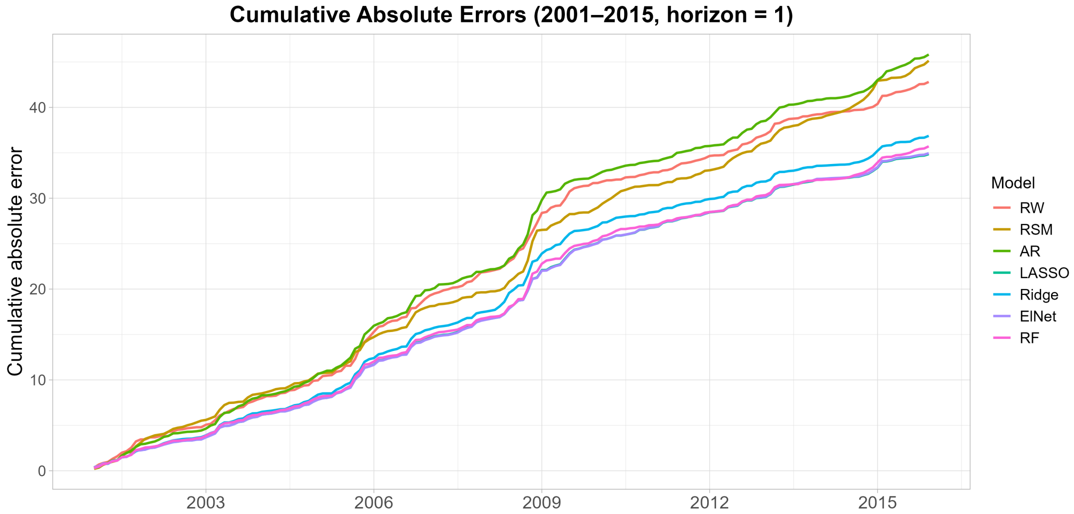


Figure 11: Cumulative absolute forecast errors for 1-month ahead predictions (2001–2015)

Note: The figure displays the cumulative sum of absolute forecast errors over time for all models during the first out-of-sample period. Lower cumulative errors indicate better overall forecast performance.

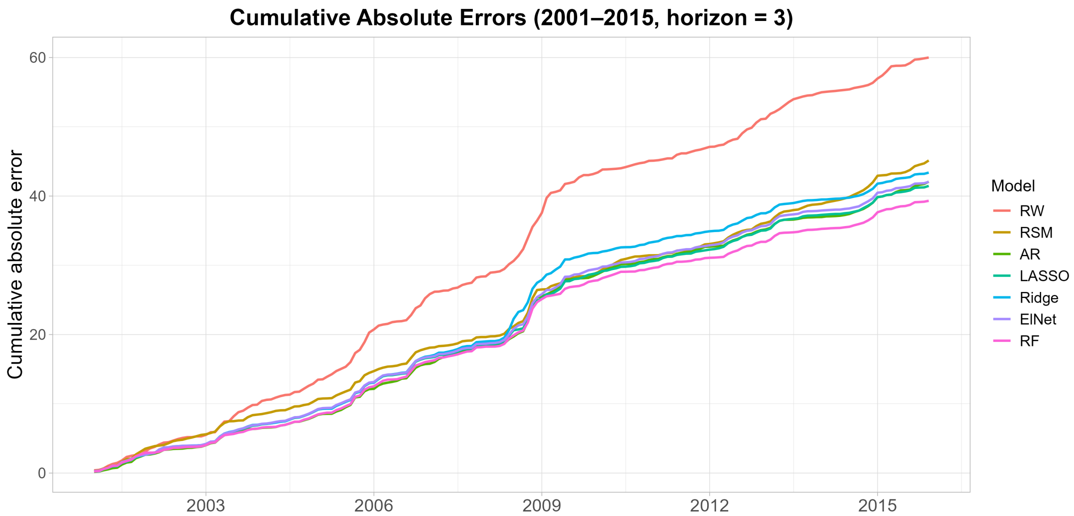


Figure 12: Cumulative absolute forecast errors for 3-month ahead predictions (2001–2015)

Note: The figure displays the cumulative sum of absolute forecast errors over time for all models during the first out-of-sample period. Lower cumulative errors indicate better overall forecast performance.

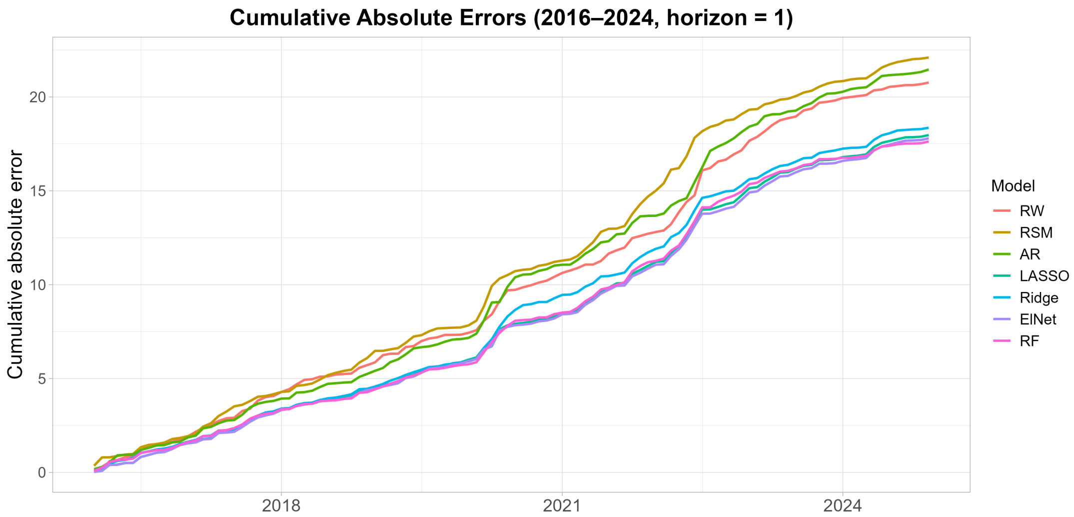


Figure 13: Cumulative absolute forecast errors for 1-month ahead predictions (2016–2024)

Note: The figure displays the cumulative sum of absolute forecast errors over time for all models during the first out-of-sample period. Lower cumulative errors indicate better overall forecast performance.

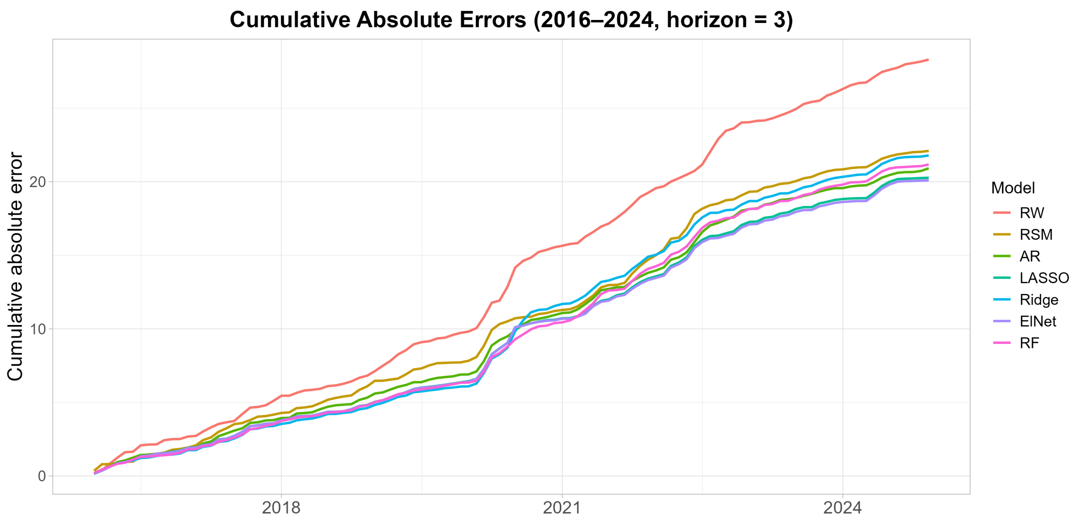


Figure 14: Cumulative absolute forecast errors for 3-month ahead predictions (2016–2024)

Note: The figure displays the cumulative sum of absolute forecast errors over time for all models during the first out-of-sample period. Lower cumulative errors indicate better overall forecast performance.

5 Conclusion

This study demonstrates that machine learning methods offer substantial improvements over traditional benchmarks for forecasting US inflation in data-rich environments. Using the FRED-MD database with 127 macroeconomic variables, we find that both regularization techniques (LASSO, Ridge, Elastic Net) and nonlinear methods (Random Forest) consistently outperform naive benchmarks across multiple forecast horizons and out-of-sample periods.

Random Forest emerges as the superior method for medium-term forecasts (3-step ahead), achieving RMSE improvements of 20–30% over the AR(4) benchmark. The model's ability to capture nonlinear relationships among employment, housing, and interest rate variables proves particularly valuable at longer horizons. For short-term forecasts (1-step ahead), regularization methods demonstrate competitive performance, with Elastic Net achieving the lowest RMSE of 0.224 in the recent period through effective variable selection from the high-dimensional feature space.

Our successful replication of Medeiros et al. (2021) for the 2001–2015 period validates our methodology, while the extended analysis through 2024 confirms that the machine learning advantage persists in recent data. The improved forecast accuracy across all models in the 2015–2025 period, despite including the post-2020 inflation surge, suggests that data-rich approaches combined with flexible modeling techniques remain valuable for inflation forecasting even in evolving economic environments.

These findings have important implications for central banks and policymakers. The combination of comprehensive datasets with machine learning methods provides a practical framework for generating more accurate inflation forecasts, which are crucial for timely and appropriate monetary policy responses.

References

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.

Joseph, A., Potjagailo, G., Chakraborty, C., and Kapetanios, G. (2024). Forecasting UK inflation bottom up. *International Journal of Forecasting*, 40(4):1521–1538.

Lenza, M., Moutachaker, I., and Paredes, J. (2025). Density forecasts of inflation: A quantile regression forest approach. *European Economic Review*, 178.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

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