Interpretable Machine Learning for Price Index Forecasting: A Case Study with Rolling Windows and SHAP

Ali Ben Mrad
Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
a.benmrad@qu.edu.sa (corresponding author)

Hamad Mohammad Alsowayyan
Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
411109618@qu.edu.sa

Received: 18 September 2025 | Revised: 13 October 2025 and 30 October 2025 | Accepted: 10 November 2025

 DOI: https://doi.org/10.48084/etasr.14863

ABSTRACT

Economic forecasting of indicators, such as the Price Index, remains a challenging task due to the strong non-stationarity and nonlinearities inherent in financial and macroeconomic time series. This study uses Portugal as a case study and compares several Machine Learning (ML) models, including Linear Regression, Ridge, Lasso, Support Vector Regression (SVR), Random Forests, Multi-Layer Perceptrons (MLPs), and XGBoost, under two evaluation protocols: a traditional static split and a rolling-window approach. The results show that static evaluation leads to poor predictive performance, with R^2 values close to zero or negative, highlighting the limitations of conventional validations when dealing with non-stationary data. In contrast, the rolling-window approach over a five-year horizon significantly improves predictive accuracy, with XGBoost achieving R^2 values above 0.94, thus confirming the adaptability of this method. To ensure interpretability, the study applied SHapley Additive Explanations (SHAP) to analyze both local and global feature contributions. The findings underline the predominant role of exchange rate, inflation, and oil prices in explaining the Portuguese Price Index. Combining rolling-window learning with explainable AI (XAI), thus, provides a robust and interpretable framework for policymakers and investors to better understand economic dynamics.

Keywords-price index forecasting; machine learning; XGBoost; explainable artificial intelligence; interpretable machine learning; rolling-window evaluation; economic policy forecasting

I. INTRODUCTION

Forecasting economic indicators, such as the Price Index, plays a crucial role for policymakers, investors, and macroeconomic analysts. However, economic time series often exhibit strong non-stationarity, meaning that their statistical properties (mean, variance, dependence structure) change over time, which complicates their modeling using traditional approaches such as ARIMA or simple linear models [1]. To address this challenge, adaptive and nonlinear ML models, combined with explainability techniques, are gaining increasing popularity in the literature.

For example, authors in [2] proposed methods to adapt ML models to non-stationary data without prior knowledge of which variables are non-stationary. Similarly, authors in [3] compared different validation techniques for non-stationary series, highlighting that forward-validation and rolling-window schemes provide better-calibrated estimates of out-of-sample performance than conventional cross-validation. Moreover, explainability approaches, such as SHAP, [4] make it possible to understand both locally and globally how explanatory variables influence the prediction of a complex model.

In this context, the present study investigates how an ML model (notably XGBoost) trained under a rolling-window protocol on data from a single country, Portugal, can provide much better predictive performance than traditional static approaches, while remaining interpretable via SHAP. The study first compares the performances of different models (linear regression, regularized models, random forests, neural networks) using a static split (80/20), then with a 5-year rolling window and a monthly step. Finally, a SHAP-based analysis is provided to explain the contributions of macroeconomic variables to local predictions.

The choice of Portugal as a case study is motivated by the availability of a long, reliable series of economic data (2002–2022), as well as by its relevance as part of the European economies, where monetary policy and inflation shocks remain central issues. While this work focuses on a single country, it paves the way for a broader multi-country extension, enabling a comparative discussion of the performance and explanatory factors across national contexts.

The investigation is conducted around the following research questions:

- RQ1. To what extent can ML models accurately predict the Price Index of a given country?
- RQ2. How does the use of rolling windows affect the stability and performance of predictive models?
- RQ3. Which features are most influential in shaping the Price Index, and how can XAI methods, such as SHAP, provide interpretable insights for financial decision-making?

In summary, this study makes three main contributions.

First, it compares several interpretable ML algorithms for Price Index forecasting under both static and rolling-window evaluation schemes. Second, it integrates SHAP-based explainability to identify the most influential macroeconomic drivers shaping price dynamics. Third, it provides empirical evidence that rolling-window learning yields superior and more stable performance on non-stationary economic data.

II. RELATED WORK

A. Price Index and Economic Time Series Prediction

The prediction of the Price Index and related economic time series has long been a central topic in econometrics and finance. Classical approaches, such as ARIMA and its variants, have been widely used to capture linear dependencies and short-term dynamics [1, 5]. GARCH models were introduced to account for volatility clustering, which is common in financial markets [6, 7]. More recent studies have explored hybrid approaches combining ARIMA or GARCH with ML models to better capture nonlinearities [8, 9]. Applications across different countries highlight the sensitivity of Price Index forecasting to structural and macroeconomic factors, emphasizing the importance of context-specific modeling [10, 11]. Research in applied forecasting has highlighted the value of hybrid and data-driven approaches in improving accuracy.

For instance, authors in [12] highlighted the capacity of advanced ML and DL techniques to improve forecasting accuracy for volatile financial data. Similarly, another study [13] demonstrated how tree-based ensemble approaches can be effectively applied to price prediction tasks when macroeconomic indicators, such as inflation, are incorporated. These studies confirm that combining classical economic determinants with modern ML techniques can substantially enhance predictive performance.

B. ML in Financial and Economic Forecasting

ML has become a powerful tool for forecasting financial and economic indicators. Algorithms, such as Random Forests, Gradient Boosting Machines, SVR, and Neural Networks have been extensively tested in predicting stock prices, exchange rates, and macroeconomic indicators [14-16]. XGBoost, in particular, has demonstrated strong performance due to its ability to handle nonlinear relationships and high-dimensional data [17]. Furthermore, deep learning models, such as LSTM networks, have been widely applied to capture temporal dependencies in time series [18-19]. Validation techniques adapted to time series, such as rolling or expanding windows, have been proposed to ensure robust evaluation [3, 20].

ML algorithms have also been widely tested in financial prediction tasks. For example, authors in [21] evaluated a range of models, including SVM, boosting algorithms, and ensemble methods, and showed their robustness in capturing nonlinear and dynamic relationships in exchange rate forecasting. These findings reinforce the importance of kernel-based and ensemble approaches for handling the complexity of economic and financial data.

C. XAI in Finance and Economics

While ML methods provide predictive power, they often suffer from a lack of interpretability, which limits their adoption in sensitive domains such as finance. XAI techniques aim to bridge this gap by providing post-hoc or model-intrinsic explanations [22]. Among these, SHAP has gained popularity as a unified framework for feature attribution [4]. Other methods, such as LIME [23], Integrated Gradients [24], and Anchors [25], have also been applied in financial forecasting tasks. In finance, XAI has been used for credit scoring [26], risk management [27], and algorithmic trading [28], demonstrating its potential to increase trust and accountability in AI-driven decision making.

Moreover, research has explored the application of ML to highly volatile assets, offering insights into model resilience under extreme conditions. In particular, authors in [29] investigated the predictive power of deep learning methods for cryptocurrencies during the COVID-19 crisis, illustrating the challenges and opportunities of forecasting in turbulent markets. This underscores the necessity of interpretable and reliable models for supporting decision-making in uncertain macroeconomic environments.

III. METHODOLOGY

A. Dataset and Features

The dataset consists of daily observations of the Portuguese Price Index alongside eight key macroeconomic variables from January 2002 to August 2021 (5033 observations) [32]. Days when stock markets were closed due to holidays or political events were removed to maintain alignment across all series. The explanatory variables include exchange rate (EUR/USD), inflation (CPI), policy interest rate, unemployment, GDP, gold prices, and oil prices. These variables are determinants of macroeconomic and financial dynamics [1], [2]. Table I summarizes the explanatory variables used in this study.

Some macroeconomic indicators, such as GDP, unemployment rate, and inflation, were available at a monthly frequency. To align them with the daily series (e.g., oil price, exchange rate, gold price, and consumer price index), a frequency harmonization procedure was applied using a monthly forward-fill interpolation. In this method, each monthly value was extended over all corresponding daily observations within the same month. This step ensures temporal consistency and avoids the introduction of artificial variability while maintaining the interpretability of economic trends.

All lagged and rolling statistics are computed using information available up to each time point only. No future observations are used when forming inputs, preserving the temporal ordering required for forecasting evaluation.

TABLE I. EXPLANATORY VARIABLES AND THEIR DESCRIPTIONS

Variable	Description
Exchange rate (EUR/USD)	Nominal exchange rate refers to USD.
Inflation rate (CPI)	Consumer Price Index (CPI), a key indicator reflecting overall economic conditions and inflationary pressures.
Interest rate (IR)	The policy interest rate set by the central bank to influence monetary variables.
Gross Domestic Product (GDP)	GDP is calculated using the expenditure method (consumption + investment + government expenditure + exports – imports).
Unemployment (UR)	The percentage of the labor force involves the currently unemployed and actively seeking employment individuals.
Gold price (G)	World gold price index (XAU/USD).
Oil price (O)	Global benchmark Brent crude oil price (USD).

B. ML Models

In this study, a diverse set of regression models was selected that represent different families of ML approaches. This variety ensures a balanced comparison between classical econometric-inspired methods, ensemble learning strategies, and modern neural architectures. The selected models are: Linear Regression [3], Ridge Regression [4], Lasso Regression [5], SVR [6], Random Forest [7], XGBoost [8], and MLP [9].

By combining linear baselines, regularized models, kernel-based methods, ensemble techniques, and neural networks, the study aims to obtain a comprehensive picture of the predictive performance and interpretability trade-offs across different learning paradigms.

While deep learning and hybrid models, such as LSTM and ARIMA–ML combinations, have shown strong predictive power in economic forecasting, they were deliberately excluded from this study. These models typically require very long-time sequences and large parameter spaces, which are not well-suited to the relatively short and heterogeneous macroeconomic series analyzed in the current work. Moreover, their "black-box" nature limits interpretability, which is a core objective of this study. Instead, focus was placed on regression and ensemble methods, such as XGBoost and Random Forest, which provide an optimal balance between accuracy, computational efficiency, and explainability. Nevertheless, integrating deep or hybrid architectures in a transparent and interpretable way remains an interesting direction for future research.

C. Evaluation Protocol and Metrics

Two protocols were used. First, a static 80/20 temporal split where the first 80% of data is used for training and the last 20% for testing. Second, a rolling-window protocol (5 years ≈ 1250 days, monthly step = 20 days), where models are trained on a moving window and tested one step ahead, enabling adaptation to non-stationary dynamics [10].

Performance was assessed with Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). RMSE penalizes large deviations, MAE provides scale-consistent accuracy, and R^2 measures the proportion of variance explained.

D. XAI Framework

Beyond predictive performance, interpretability was addressed with SHAP [11]. This method attributes the contribution of each explanatory variable to individual predictions (local interpretability) while also enabling global insights into the most influential drivers across the dataset. By combining rolling-window modeling with SHAP analysis, the study ensures both robust predictive accuracy and transparent explanations of the role of macroeconomic factors.

The pseudocodes illustrating the main algorithmic steps (frequency alignment, rolling evaluation, and SHAP computation) are:

```
1) Rolling-Window Evaluation Protocol
Input: dataset D with features X(t),
target y(t)
Set window_length = W, step_size = s
For t = 1 to T - W step s:
    Train model M_t on [t, t + W - 1]
    Evaluate on [t + W, t + W + s - 1]
    Store metrics: RMSE_t, MAE_t, R²_t
Compute mean and std of metrics across all
windows
2) Frequency Alignment Procedure
For each monthly indicator series S_m:
    For each day d in month m:
        S_d = S_m  # forward-fill within
month
Merge all daily series by date key
3) SHAP-Based Explainability (Global
Aggregation Across Windows)
Input: trained models {M_t} from rolling
protocol, evaluation sets {X_eval,t}
Initialize aggregator A[f] = [] for each
feature f
For each window t:
    For each observation x_i in X_eval,t:
        SHAP_i = shap_values(M_t, x_i)
        For each feature f:
            Append |SHAP_i[f]| to A[f]
For each feature f:
    Compute GlobalImportance[f] =
mean(A[f])
Rank features by GlobalImportance
4) Single-Instance SHAP Explanation
Example
Input: trained model M_{t*}, single
observation x* at time t*
Baseline: define reference distribution
```

```
from training window [t* - W, t* - 1]
Compute SHAP* = shap_values(M_{t*}, x*)
Sort features by |SHAP*|
Return Top-k pairs (feature, SHAP
contribution) to explain y_hat(x*)
```

IV. RESULTS AND ANALYSIS

A. Descriptive Analysis of the Time Series

Table II provides descriptive statistics of the explanatory variables and the Portuguese Price Index between 2002 and 2021. The statistics confirm significant variability across variables such as oil and gold, while macroeconomic variables, like GDP, show relatively stable dynamics. The Price Index itself exhibits high volatility, with a skewed distribution and high kurtosis, highlighting the challenge of modeling.

Figure 1 illustrates the temporal evolution of the Portuguese Price Index from 2002 to 2022. The series shows strong fluctuations, with a sharp peak before the 2008 financial crisis, followed by a significant decline and subsequent periods of partial recovery. These dynamics highlight the non-stationary nature of the series, justifying the use of adaptive models for forecasting.

TABLE II. DESCRIPTIVE STATISTICS OF EXPLANATORY VARIABLES AND PRICE INDEX (PORTUGAL, 2002–2021)

Variable	Mean	Std	Min	25%	50%	75%	Max
Exchange rate (EUR/USD)	1.23	0.13	0.85	1.13	1.22	1.33	1.59
Inflation rate	101.53	8.15	83.44	95.27	104.88	107.91	112.22
Interest rate	1.28	1.33	0.00	0.00	1.00	2.00	4.25
Unemployment	10.20	3.30	5.32	7.45	8.98	12.35	18.22
GDP	4.43e10	4.84e9	3.53e10	4.17e10	4.40e10	4.69e10	5.43e10
Gold	1073.95	467.16	278.20	641.50	1203.39	1355.69	2063.18
Oil	67.88	28.02	18.41	47.63	63.69	86.10	146.08
Price Index	6731.80	2068.76	3596.08	5233.86	6026.08	7665.53	13702.0

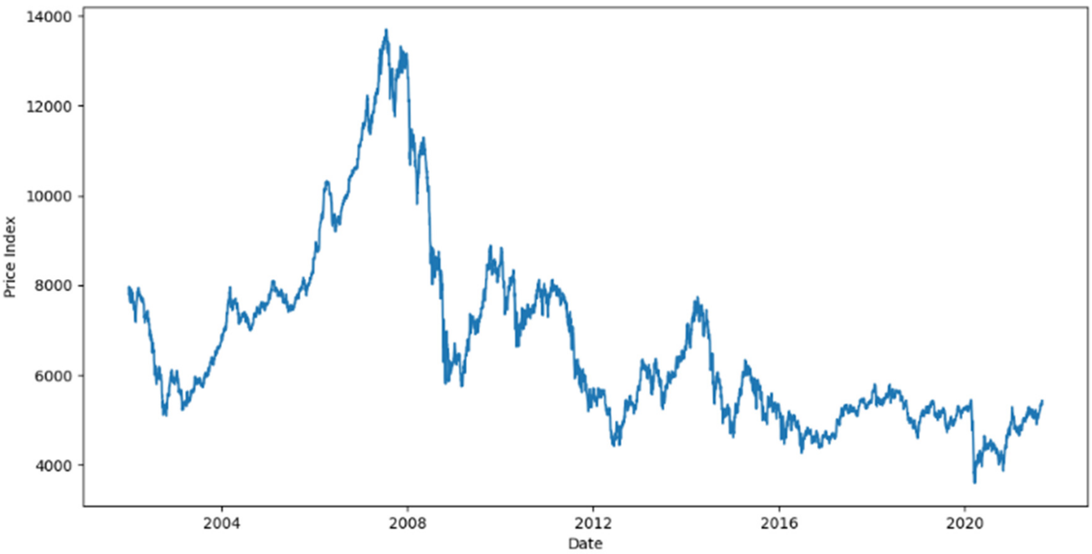

Fig. 1. Evolution of the Portuguese Price Index (January 2002 - August 2021).

B. Correlation Analysis

To better understand the linear relationships between the explanatory variables and the Portuguese Price Index, the Pearson correlation coefficients were computed. These results highlight which variables are most strongly linked to the index dynamics, whether positively or negatively. Figure 2 illustrates these correlations. The Interest Rate shows the highest positive correlation with the Price Index (≈ 0.78), followed by the Exchange Rate (EUR/USD) (≈ 0.52) and Oil prices (≈ 0.18). These findings suggest that monetary conditions and international market fluctuations play a decisive role in shaping the Portuguese Price Index.

The Interest Rate shows the highest positive correlation (≈ 0.78), followed by the Exchange Rate (EUR/USD) (≈ 0.52) and Oil (≈ 0.18). Gold and the Inflation Rate exhibit negative correlations (≈ –0.47 and –0.46, respectively). On the other hand, some variables display a negative correlation. Gold and the Inflation Rate present correlations of approximately -0.47 and -0.46, respectively, reflecting their role as counter-indicators in the studied context. GDP and Unemployment appear weakly correlated, likely due to the complexity of macroeconomic dynamics and the time lag between these variables and index movements. This negative relationship between unemployment and the Price Index is consistent with the Phillips curve framework, which suggests that higher unemployment tends to reduce inflationary pressures, while lower unemployment may coincide with rising prices. These descriptive correlations represent a first step in the analysis. While they do not establish causal relationships, they provide useful insights into which variables may exert significant influence in the predictive models tested later.

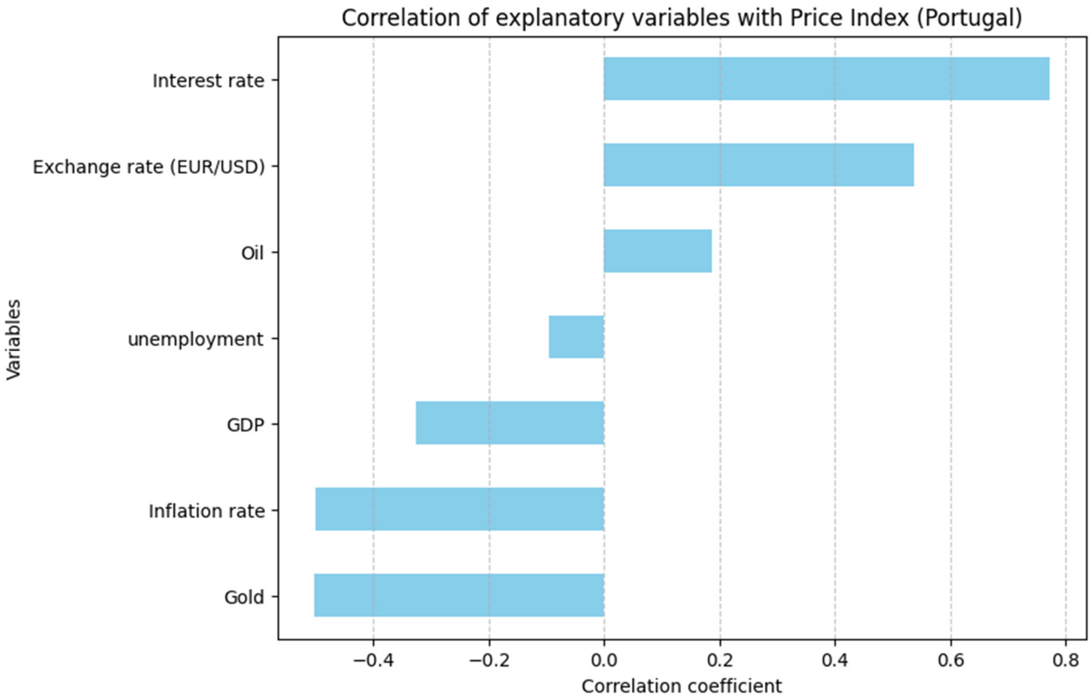

Fig. 2. Correlation of explanatory variables with the Price Index (Portugal).

C. Model Performance: Static Split (80/20)

In the first stage, the performance of regression models was evaluated using a conventional static split of the dataset (80% for training and 20% for testing). This methodology is commonly adopted to provide a baseline estimate of the models' generalization ability.

The results reported in Table III highlight the limitations of static evaluation. The best-performing model, XGBoost, achieved an RMSE of 434.09, a MAE of 308.82, and an R^2 = 0.036, indicating a very limited explanatory power. Random Forest obtained a similarly poor R^2 close to zero (0.005). Linear models (Ridge, Lasso, Linear Regression) performed even worse, with negative R^2 values, reflecting a failure to capture the underlying dynamics. Finally, the MLP displayed the weakest performance, with R^2 = -23.40, underscoring instability and overfitting.

TABLE III. STATIC SPLIT MODEL PERFORMANCE

Model	RMSE	MAE	R^2
XGBoost	434.09	308.82	0.036
RandomForest	440.86	331.28	0.005
SVR	893.90	816.26	-3.089
Ridge	1849.43	1313.53	-16.505
Lasso	1869.99	1326.58	-16.897
LinearRegression	1870.39	1326.83	-16.904
MLP	2183.64	2040.45	-23.403

These results suggest that the static split is inadequate for modeling non-stationary financial time series. They motivate the use of more appropriate evaluation protocols, such as rolling-window approaches, which are explored below.

D. Model Performance: Rolling Evaluation

The rolling window evaluation provides a more realistic framework for assessing model performance on non-stationary economic time series. Unlike the static 80/20 split, this method continuously updates the training set, reflecting real-world forecasting conditions where models must adapt over time.

TABLE IV. ROLLING WINDOW EVALUATION RESULTS (WINDOW = 1250 DAYS, STEP = 20)

Model	RMSE	MAE	R^2
XGBoost	520.51	151.91	0.943
RandomForest	588.10	167.68	0.928
SVR	810.19	407.59	0.863
LinearRegression	820.77	526.08	0.859
Lasso	820.80	526.07	0.859
Ridge	830.23	527.24	0.856
MLP	1941.71	1476.03	0.212

As shown in Table IV, all models perform significantly better under rolling evaluation compared to static splitting. XGBoost achieves the best results with an R^2 of 0.94, followed by Random Forest ($R^2 \approx 0.93$). SVR and linear models also demonstrate strong performance, with R^2 values between 0.85 and 0.86. In contrast, the MLP underperforms, showing difficulties in capturing temporal dynamics with limited training windows. The weak performance of the MLP model can be explained by its higher sensitivity to feature scaling and the relatively small dataset size compared to its parameter complexity. MLPs generally require extensive tuning and large training samples to perform well. In contrast, ensemble models, such as XGBoost and Random Forest, are more robust for structured macroeconomic data, providing strong predictive accuracy with limited preprocessing requirements.

Figure 3 illustrates the comparison between the true Price Index values and the predictions generated by the XGBoost model under the rolling evaluation scheme. The close alignment of the two curves highlights the strong predictive power and stability of the model.

To prevent potential overfitting in the rolling-window evaluation, early stopping and a temporal validation split were applied during XGBoost training. Hyperparameters, such as learning rate, depth, and the number of trees, were tuned to balance model complexity and predictive accuracy. Additional experiments using time-series cross-validation and early stopping confirmed the robustness of the rolling-window approach. The static validation setup showed unstable and often negative R^2 values, highlighting its inadequacy for non-stationary data. In contrast, the rolling evaluation with regularization achieved consistent and realistic results (R^2 = 0.949 for XGBoost), demonstrating that this dynamic validation protocol better reflects real-world forecasting conditions.

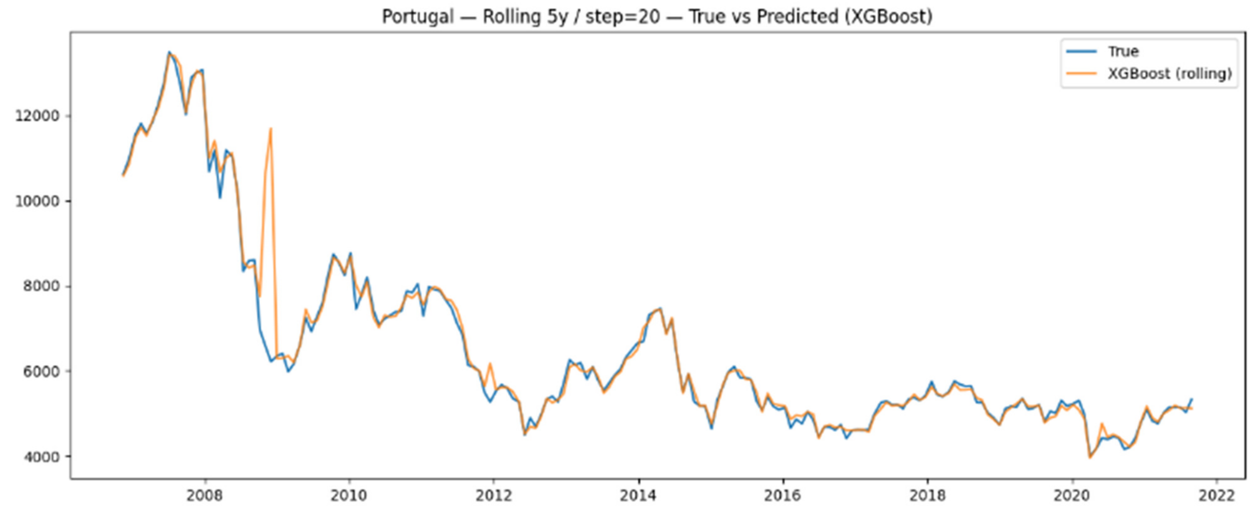

Fig. 3. True versus predicted values of the Price Index using XGBoost with a rolling window evaluation.

E. Rolling Evaluation Diagnostics

Performance is reported over multiple rolling windows under a consistent protocol. The results are stable across windows without systematic deviations between training and test performance. The relatively high out-of-sample R^2 reflects short-horizon adaptability in a time-updated window and should not be construed as evidence of long-range predictability.

F. Explainability with SHAP

To complement predictive performance evaluation, the SHAP was applied to interpret the XGBoost model trained under the rolling window framework. SHAP provides both global and local explanations, highlighting the relative contribution of each macroeconomic variable to the predicted Price Index. Figures 4 and 5 depict the key explainability results. Figure 4 presents the mean absolute SHAP values across features. Oil emerges as the most influential predictor, followed by unemployment, gold, and inflation. Exchange rate, GDP, and interest rate play a comparatively minor role. This ranking provides a global understanding of the main drivers of the Portuguese Price Index in the considered period. Oil (≈ 230) and Unemployment (≈ 76) are the most influential predictors, followed by Gold (≈ 46) and Inflation (≈ 36). Exchange Rate, GDP, and Interest Rate have lower contributions. Figure 5 shows the SHAP summary plot, which combines local and global perspectives. Each dot represents a single observation, colored by the feature value (blue = low, red = high). This helps understand not only which features are important, but also how their effects vary depending on their magnitude. For example, high oil prices (red points) consistently contribute positively to the price index, confirming the inflationary impact of energy costs. In contrast, high unemployment (red points) is systematically associated with negative contributions, indicating its downward pressure on the index. Inflation exhibits a more heterogeneous effect: in some cases, high inflation increases the predicted index sharply, while in others the effect is moderate, reflecting possible periods of monetary policy intervention. For GDP and exchange rate, the points are more concentrated around zero, highlighting their relatively stable and secondary influence.

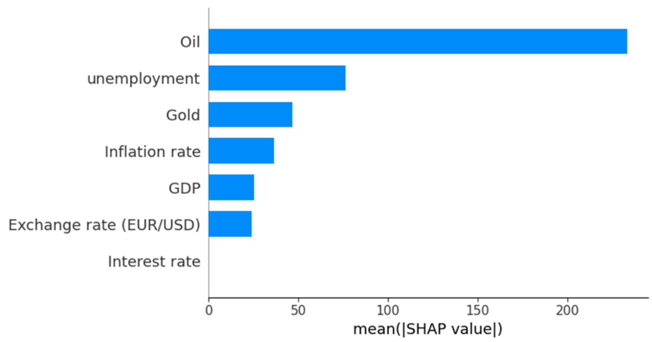

Fig. 4. Feature importance based on mean absolute SHAP values.

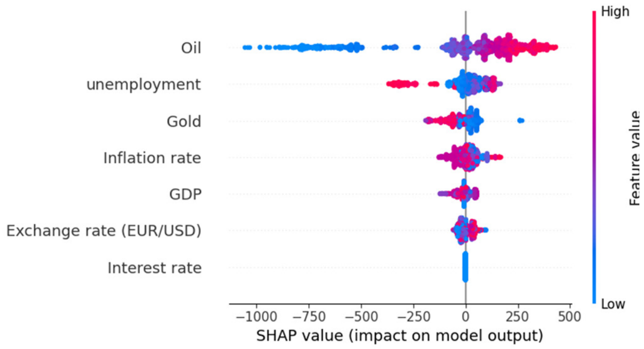

Fig. 5. SHAP summary plot with feature value coloring.

Overall, the SHAP analysis demonstrates that oil and unemployment are the dominant drivers of the Portuguese Price Index, while other variables play more nuanced roles. This explainability perspective enhances transparency and trust in the proposed ML approach. Beyond identifying the most influential variables, SHAP analysis offers practical implications for policy and decision-making. Quantifying the marginal impact of each macroeconomic factor on price dynamics enables policymakers to monitor the sources of inflationary pressure in real time. For example, positive SHAP values associated with oil prices or unemployment may indicate cost-push inflation, while negative contributions from GDP suggest economic deceleration. Such transparent interpretability bridges the gap between data-driven forecasting and actionable economic policy.

The strong correlation observed between the interest rate and the Price Index does not necessarily translate into a high SHAP importance. Correlation reflects a global linear relationship, while SHAP values measure the conditional and marginal contribution of each variable within the model. Once nonlinear dependencies with oil, inflation, and unemployment are considered, the incremental predictive power of the interest rate becomes relatively small, explaining its lower SHAP importance.

To complement the global analysis, Figure 6 presents a local SHAP explanation for a specific observation corresponding to a period of increased oil prices.

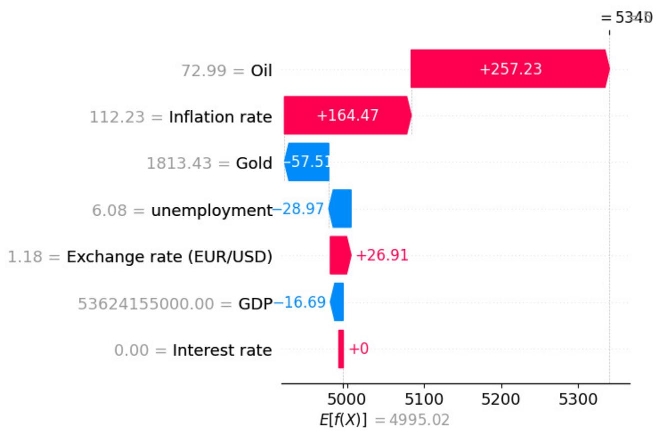

Fig. 6. Local SHAP explanation (waterfall plot).

The waterfall plot decomposes the prediction into additive feature contributions relative to the model's base value. The results indicate that oil and inflation rate exerted the largest positive influence on the predicted Price Index, while unemployment and GDP had a dampening effect. This local interpretation highlights the model's ability to capture the causal intuition behind macroeconomic fluctuations, notably, how oil shocks tend to raise consumer prices through higher production and transportation costs.

V. CONCLUSION AND FUTURE WORK

This study has shown that forecasting the Price Index can be substantially improved by combining modern Machine Learning (ML) models with a rolling-window evaluation and training scheme. Whereas a conventional static 80/20 split yielded limited accuracy, the rolling approach revealed that tree-based ensembles, most notably XGBoost, achieve very high predictive performance, with out-of-sample coefficients of determination exceeding 0.9. In practical terms, rolling windows are a more suitable validation and deployment strategy for non-stationary economic time series than conventional random or chronological hold-outs, because they continuously adapt the training data to the most recent regime.

Beyond accuracy, the study emphasized interpretability through explainable Artificial Intelligence (XAI). Using SHapley Additive Explanations (SHAP), both global and local contributions of macroeconomic drivers were quantified to model predictions. The analysis highlighted oil prices and unemployment as dominant factors for the Portuguese Price Index over the study period, while inflation, GDP, the exchange rate, and the policy rate exhibited more nuanced and context-dependent effects. Such explanations increase transparency and trust in data-driven forecasts, which is essential in policy-relevant economic applications.

The proposed framework demonstrates that combining predictive accuracy with interpretability enhances both analytical robustness and practical usability. By combining high-performing models with transparent SHAP-based explanations, the approach supports trustworthy and evidence-based decision-making in economic forecasting.

The study has limitations. The experiments focused on a single country (Portugal) and a single target series. Consequently, external validity should be considered with caution, as structural characteristics and policy frameworks differ across countries and over time. To address this limitation, future research will aim to extend the proposed framework to additional European countries, such as Spain, Italy, and France, allowing for cross-country benchmarking and sensitivity analysis. This comparative evaluation will test the external validity of the interpretable forecasting approach across different macroeconomic contexts.

Future work will pursue three directions. First, a multi-country extension will enable systematic comparisons of predictive performance and feature attribution across heterogeneous economic environments. Second, incorporating additional information, such as news or social media-based sentiment indicators, commodity supply shocks, or climate-related variables, may capture exogenous drivers that are only partially reflected in macro-aggregates. Third, an operational study of rolling-window deployment (choice of window length and update frequency) can help practitioners balance responsiveness to regime shifts with computational cost.

Research has further highlighted the growing role of interpretable ML in economic forecasting. For instance, authors in [30] combined SHAP-based interpretability with news sentiment indicators to enhance the prediction of economic cycles, while authors in [31] demonstrated how combining multiple large datasets with ML methods improves macroeconomic forecasting accuracy. These developments reinforce the importance of combining transparent interpretability techniques with adaptive learning strategies, as explored in this work. This alignment with recent developments underscores the timeliness and relevance of the proposed framework.

Traditional econometric baselines (e.g., ARIMA, VAR, and a naïve Random Walk) are acknowledged as pertinent benchmarks, and can be incorporated under the same rolling protocol in future work.

Overall, the results demonstrate the value of hybrid approaches that integrate robust ML, time-aware evaluation protocols, and principled explainability to support the understanding and forecasting of complex macroeconomic phenomena.

ACKNOWLEDGMENT

The authors gratefully acknowledge Qassim University, represented by the Deanship of Graduate Studies and Scientific Research, on the financial support for this research under the number (QU-J-UG-2-2025-56270) during the academic year 1446 AH/2024 AD.

DATA AVAILABILITY STATEMENT

The dataset utilized in this study, the Price Index Dataset, is publicly available and can be downloaded from: https://github.com/AliBenMrad/price-index-dataset/raw/refs/heads/main/BD%20Portugal%20NS.xlsx [32].

REFERENCES

[1] R. T. Baillie, "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, vol. 73, no. 1, pp. 5–59, July 1996, https://doi.org/10.1016/0304-4076(95)01732-1.

[2] J. Xi, "Machine Learning Using Nonstationary Data." SSRN Preprint, 2025, https://doi.org/10.2139/ssrn.5215521.

[3] M. Schnaubelt, "A Comparison of Machine Learning Model Validation Schemes for Non-stationary Time Series Data," arXiv, 2019, https://doi.org/10.13140/RG.2.2.29545.24168.

[4] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," arXiv, 2017, https://doi.org/10.48550/ARXIV.1705.07874.

[5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2016.

[6] R. F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, vol. 50, no. 4, July 1982, pp. 987–1007, https://doi.org/10.2307/1912773.

[7] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, Apr. 1986, https://doi.org/10.1016/0304-4076(86)90063-1.

[8] G. P. Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, https://doi.org/10.1016/S0925-2312(01)00702-0.

[9] R. S. Tsay, *Analysis of Financial Time Series*, 1st ed. Hoboken, NJ, USA: John Wiley & Sons, 2005.

[10] M. G. P. Garcia, M. C. Medeiros, and G. F. R. Vasconcelos, "Real-time Inflation Forecasting with High-dimensional Models: The Case of Brazil," *International Journal of Forecasting*, vol. 33, no. 3, pp. 679–693, July 2017, https://doi.org/10.1016/j.ijforecast.2017.02.002.

[11] H. Hassani and E. S. Silva, "Forecasting UK Consumer Price Inflation using Inflation Forecasts," *Research in Economics*, vol. 72, no. 3, pp. 367–378, Sept. 2018, https://doi.org/10.1016/j.rie.2018.07.001.

[12] M. A. Alwadi, "Fuel Sales Price Forecasting using Time Series, Machine Learning, and Deep Learning Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22360–22366, June 2025, https://doi.org/10.48084/etasr.10348.

[13] Y. Liu, Z. Li, R. Nekhili, and J. Sultan, "Forecasting cryptocurrency returns with machine learning," *Research in International Business and Finance*, vol. 64, 2023, Art. no. 101905, https://doi.org/10.1016/j.ribaf.2023.101905.

[14] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020, https://doi.org/10.1109/ACCESS.2020.3015966.

[15] T. Fischer and C. Krauss, "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, Oct. 2018, https://doi.org/10.1016/j.ejor.2017.11.054.

[16] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: Results, Findings, Conclusion and Way Forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, Oct. 2018, https://doi.org/10.1016/j.ijforecast.2018.06.001.

[17] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[18] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, https://doi.org/10.1162/neco.1997.9.8.1735.

[19] D. M. Q. Nelson, A. C. M. Pereira, and R. A. De Oliveira, "Stock Market's Price Movement Prediction with LSTM Neural Networks," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 1419–1426, https://doi.org/10.1109/IJCNN.2017.7966019.

[20] L. J. Tashman, "Out-of-Sample Tests of Forecasting Accuracy: an Analysis and Review," *International Journal of Forecasting*, vol. 16, no. 4, pp. 437–450, Oct. 2000, https://doi.org/10.1016/S0169-2070(00)00065-0.

[21] M. El Mahjouby, K. El Fahssi, M. Taj Bennani, M. Lamrini, and M. El Far, "Machine Learning Techniques for Predicting and Classifying Exchange Rates between US Dollars and Japanese Yen," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16266–16271, Oct. 2024, https://doi.org/10.48084/etasr.8216.

[22] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," arXiv, 2017, https://doi.org/10.48550/ARXIV.1702.08608.

[23] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA, Aug. 2016, pp. 1135–1144, https://doi.org/10.1145/2939672.2939778.

[24] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," in *34th International Conference on Machine Learning*, Sydney, Australia, 2017, Art. no. 70.

[25] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-Precision Model-Agnostic Explanations," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018, https://doi.org/10.1609/aaai.v32i1.11491.

[26] D. Martens, J. Vanthienen, W. Verbeke, and B. Baesens, "Performance of Classification Models from a User Perspective," *Decision Support Systems*, vol. 51, no. 4, pp. 782–793, Nov. 2011, https://doi.org/10.1016/j.dss.2011.01.013.

[27] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining Anomalies Detected by Autoencoders using SHAPLEY Additive Explanations," *Expert Systems with Applications*, vol. 186, Dec. 2021, Art. no. 115736, https://doi.org/10.1016/j.eswa.2021.115736.

[28] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion*, vol. 58, pp. 82–115, June 2020, https://doi.org/10.1016/j.inffus.2019.12.012.

[29] N. Köse, Y. E. Gür, and E. Üna, "Deep Learning and Machine Learning Insights Into the Global Economic Drivers of the Bitcoin Price," *Journal of Forecasting*, vol. 44, no. 5, pp. 1666-1698, 2025, https://doi.org/10.1002/for.3258.

[30] W. Sun, Y. Wang, L. Zhang, X. H. Chen, and Y. H. Hoang, "Enhancing Economic Cycle Forecasting Based on Interpretable Machine Learning and News Narrative Sentiment," *Technological Forecasting and Social Change*, vol. 215, June 2025, Art. no. 124094, https://doi.org/10.1016/j.techfore.2025.124094.

[31] N. Anesti, E. Kalamara, and G. Kapetanios, "Forecasting with Machine Learning Methods and Multiple Large Datasets," *Econometrics and Statistics*, Sept. 2024, Art. no. SSRN5311094, https://doi.org/10.1016/j.ecosta.2024.08.003.

[32] A. B. Mrad, "Price Index Dataset." GitHub, 2025, [Online]. Available: https://github.com/AliBenMrad/price-index-dataset/raw/refs/heads/main/BD%20Portugal%20NS.xlsx.