How Do Macroeconomic Indicators Impact the 10-Year Treasury Yield?

Abstract

The 10-Year Treasury Yield serves as a critical benchmark in financial markets, influencing borrowing costs, investment decisions, and economic sentiment. This study explores how macroeconomic indicators such as the Federal Funds Rate, VIX, S&P 500, Breakeven Inflation Rate, and Bond Aggregate Index explain yield movements. Using machine learning models, including ARIMA, XGBoost, and LSTM, we analyzed 10 years of daily data to uncover key drivers of the yield. The LSTM model emerged as the most effective, achieving an RMSE of 0.0192 and an R^2 of 0.9487, highlighting its ability to capture both temporal and multivariate dependencies. While the LSTM demonstrated strong explanatory power, limitations such as its black-box nature and sensitivity to stable trends were identified. These findings provide insights into the complex dynamics of Treasury Yields and lay the foundation for further exploration in predictive and explanatory modeling.

Introduction

Variable Description

The independent variables used in this study capture key economic and market dynamics influencing the 10-Year Treasury Yield:

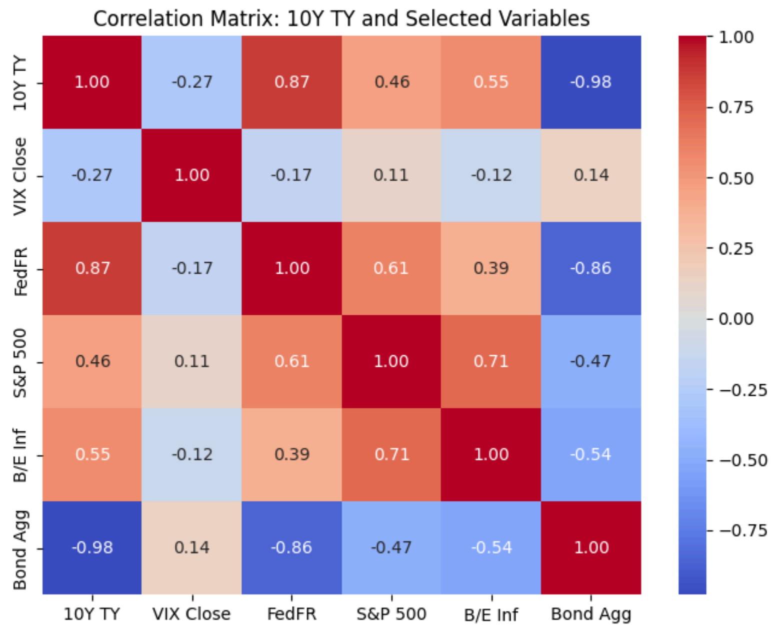
- **Fed Funds Rate**: The Federal Reserve's benchmark interest rate, a primary tool of monetary policy, influencing short-term rates and economic activity.
- **VIX Close**: The "fear gauge," reflecting market expectations of near-term volatility and investor uncertainty.
- **S&P 500 Price**: A barometer of U.S. economic performance and investor sentiment, based on the stock prices of 500 leading companies.
- **10-Year Breakeven Inflation Rate**: The difference between nominal Treasury yields and inflation-protected securities (TIPS), indicating market expectations for inflation over 10 years.
- **Bond Aggregate Index**: A measure of the U.S. investment-grade bond market, representing Treasuries, corporate bonds, and other fixed-income securities.

These variables collectively represent monetary policy, market sentiment, inflation expectations, and bond market dynamics. Using daily data from November 2014 to November 2024, they provide a comprehensive framework for modeling yield movements.

Descriptive Statistics

The correlation matrix reveals the relationships between the 10-Year Treasury Yield (10Y TY) and key macroeconomic variables. The **Fed Funds Rate** (0.87) and **Breakeven Inflation** (0.55) show positive correlations, indicating their significant role in influencing long-term yields through monetary policy and inflation expectations. In contrast, the **Bond Aggregate Index** (-0.98) has a strong negative correlation, consistent with bond price-yield dynamics.

Other variables, like the **S&P 500** (0.46) and **VIX Close** (-0.27), exhibit weaker correlations, suggesting a less direct but contextual influence on Treasury yields, primarily reflecting market sentiment and economic expectations. Overall, the matrix highlights the Fed Funds Rate, Breakeven Inflation, and Bond Aggregate Index as the primary drivers of 10Y TY, with additional context provided by S&P 500 and VIX.



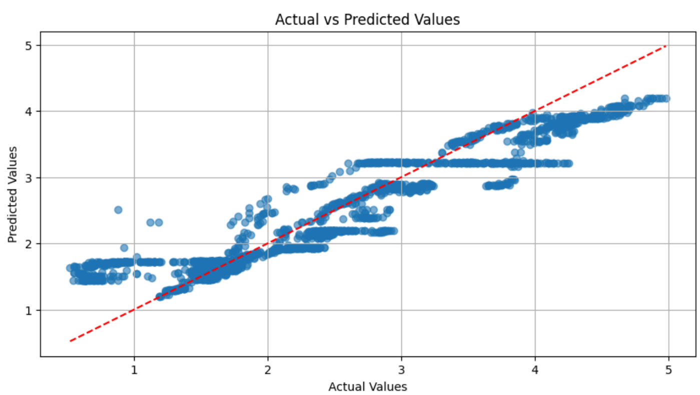
Models

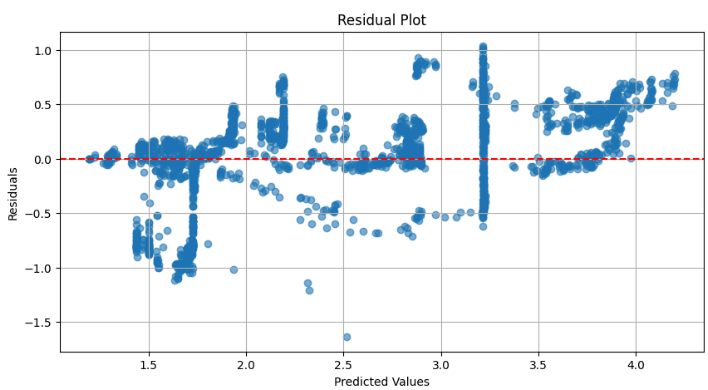
Random Forest

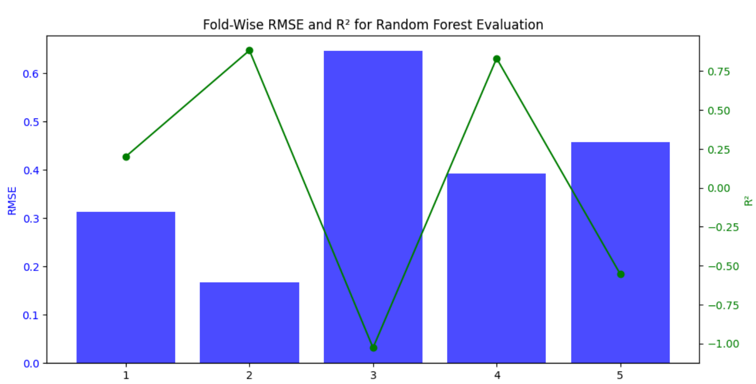
We applied Random Forest to analyze the impact of macroeconomic variables on the 10-Year Treasury Yield (10Y TY) due to its strong capability to model non-linear relationships and complex feature interactions. The algorithm's robustness to noise and flexibility in capturing varying feature relationships made it a suitable choice for our dataset. To optimize model performance, we conducted hyperparameter tuning using grid search combined with time-series cross-validation, ensuring that temporal dependencies in the data were preserved. Parameters such as the number of estimators, maximum tree depth, and split thresholds were adjusted, and RMSE and R^2 were employed as evaluation metrics to assess predictive accuracy and explanatory power.

The Random Forest model demonstrated reasonable predictive capability, as evidenced by its alignment with the ideal fit line in most cases. However, residual patterns indicated non-random errors, particularly at extreme predicted values, suggesting potential biases or unmodeled complexities in the data. This was further reflected in the residual plot, which showed higher variance at extreme values, revealing the model's limitations in generalizing across the full feature space. Additionally, fold-wise performance highlighted variability, with most folds showing satisfactory results, but one (Fold 3) significantly underperforming with a high RMSE and negative R^2 values. This inconsistency pointed to structural changes or anomalies in the data that Random Forest, as a non-temporal model, struggled to capture effectively.

Fold 3 (Train Period: 0–1258, Test Period: 1259–1677) coincided with the volatile market dynamics of November 2019 to July 2021, a period defined by the COVID-19 pandemic and its economic aftermath. During this time, the 10Y TY experienced a historic low of 0.52% and a sharp recovery to 1.94%. Similarly, the VIX peaked at 82.69, reflecting extreme uncertainty, while the Federal Funds Rate remained near-zero due to aggressive monetary easing. The S&P 500 exhibited dramatic fluctuations, from a pandemic crash low of 2,237.4 to a recovery high of 4,352.3, while bond aggregates remained relatively stable. These unprecedented conditions strained the Random Forest model's ability to generalize, highlighting its limitations in capturing temporal trends during periods of macroeconomic instability.



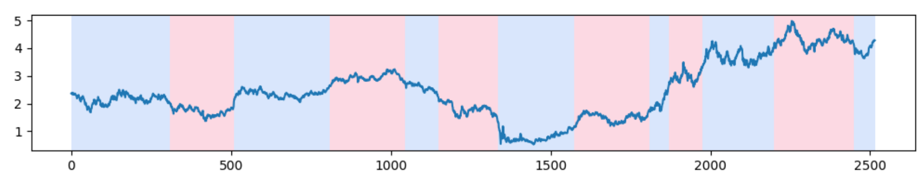


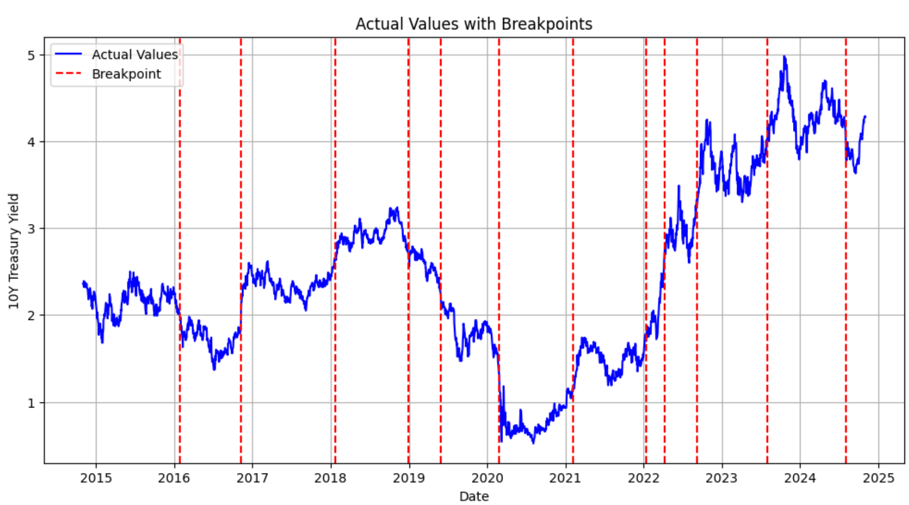


Breakpoints

The poor performance of Fold 3 highlighted structural breaks as a significant issue for the model. A CUSUM test confirmed substantial shifts in the time series' statistical properties, with a p-value of 1.0038e-268 rejecting the null hypothesis of no structural breaks. Using TimeSeriesSplit, breakpoints such as [310, 510, 810, 1045, 1150, 1335, 1575, 1810, 1870, 1975, 2200, 2450] were identified, signaling major changes in macroeconomic dynamics impacting the 10-Year Treasury Yield.

These breakpoints align with key historical events, including the aftermath of the 2016 Chinese stock market crash and U.S. election uncertainty, the 2018 market correction influenced by rate hikes and trade wars, and the COVID-19 pandemic's onset in 2020. Subsequent periods captured economic recovery amid inflation concerns, supply chain disruptions, and recent Federal Reserve rate hikes alongside geopolitical tensions. These findings underscore the importance of accounting for structural shifts to improve the model's forecasting and explanatory capabilities.



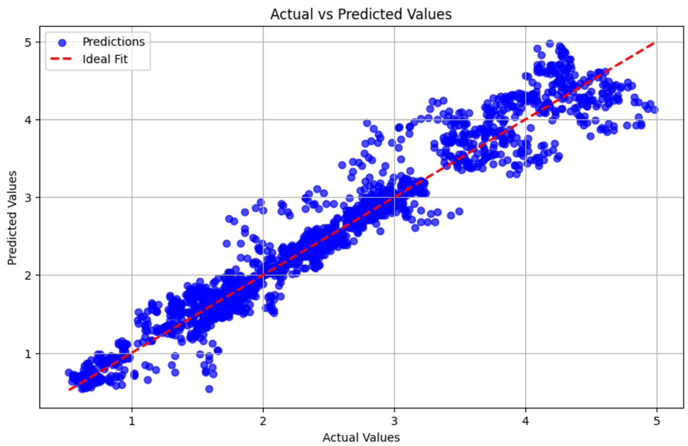


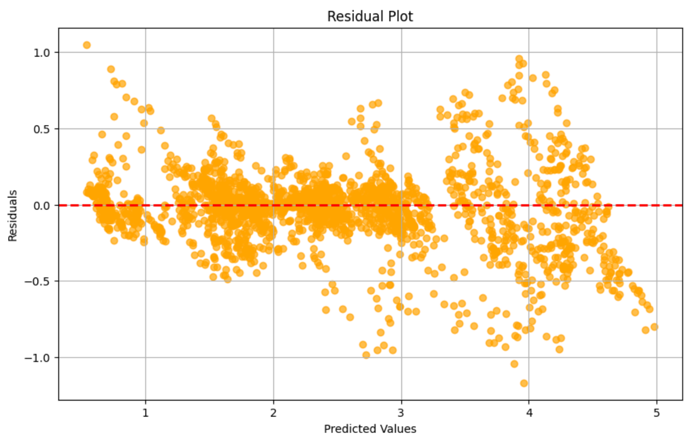
Random Forest + SARIMAX

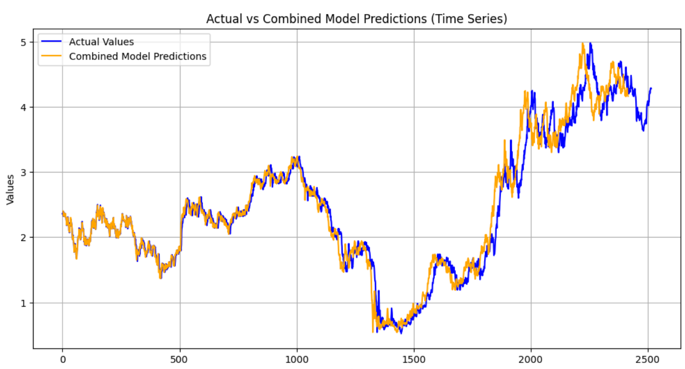
To take both structural breaks and non-linear dependencies of our time-series data, we integrated SARIMAX and Random Forest models to build a combined model. SARIMAX captures linear relationships, seasonal patterns, and structural breaks by splitting the dataset into distinct regimes based on detected breakpoints. Each regime is modeled independently, ensuring that the linear component adapts to variations in market dynamics, such as economic shifts or policy changes. Meanwhile, Random Forest complements SARIMAX by addressing residuals—the unexplained variance left by SARIMAX. By using lagged features and exogenous variables, Random Forest models complex, non-linear relationships, filling gaps that SARIMAX alone cannot address. This hybrid approach leverages the strengths of both techniques, creating a robust predictive framework.

The results provide strong evidence of the model's efficacy. The RMSE of 0.2366 highlights the model's ability to make precise predictions, while the R^2 of 0.9432 demonstrates that the model captures 94.32% of the variance in the 10-year Treasury Yield, underscoring its predictive power. The actual vs. predicted plot shows tight clustering around the ideal fit line, reflecting high accuracy and low bias. The residual plot, where residuals are evenly distributed around zero with no visible patterns, confirms that the model captures systematic variance effectively and avoids overfitting. The time-series plot shows the model's predictions closely tracking actual values, even during periods of structural change, which validates the SARIMAX component's ability to adapt to shifts in regime. The breakpoint plot, with red vertical lines marking structural shifts, illustrates how well the model segments the dataset, ensuring tailored predictions for each regime.

These results highlight important insights. First, explicitly addressing structural breaks significantly enhances the model's robustness, allowing it to adapt to regime changes that would otherwise degrade accuracy. Second, the hybrid model's ability to combine SARIMAX's linear forecasting with Random Forest's non-linear pattern recognition ensures comprehensive coverage of relationships in the dataset. Finally, the detected breakpoints align with known economic events, emphasizing the model's potential for understanding macroeconomic trends and improving decision-making in forecasting tasks. This hybrid approach demonstrates how complementary models can be integrated to create a highly effective predictive framework for complex, real-world data.


Actual vs Predicted Values
Predictions
Ideal Fit
Predicted Values
Actual Values


Residual Plot
Residuals
Predicted Values


Actual vs Combined Model Predictions (Time Series)
Actual Values
Combined Model Predictions
Values
Time

Linear regression

Linear regression establishes the relationship between the 10-Year Treasury Yield (dependent variable) and five macroeconomic indicators (independent variables) by fitting a straight line that minimizes prediction errors.

Algorithm:

1. Independent variables (X) and the dependent variable (y) are defined.
2. The data is split into training (80%) and testing (20%) sets using train_test_split.
3. The model is trained using training data..
4. Predictions are made on the test set.
5. Prediction accuracy and residuals are evaluated.
6. Feature importance is extracted via the model coefficients.

Performance Metrics:

Root Mean Squared Error (RMSE): Measures the average magnitude of prediction errors, with lower values indicating better performance.

R^2 (Coefficient of Determination): Evaluates how well the model explains the variability in the target variable. Ranges from 0 to 1, with higher values indicating better performance.

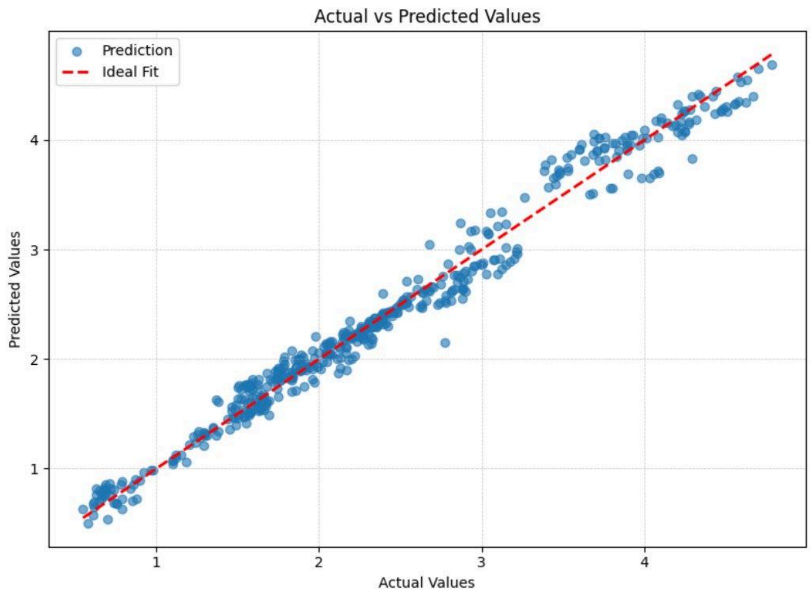
Model Performance:

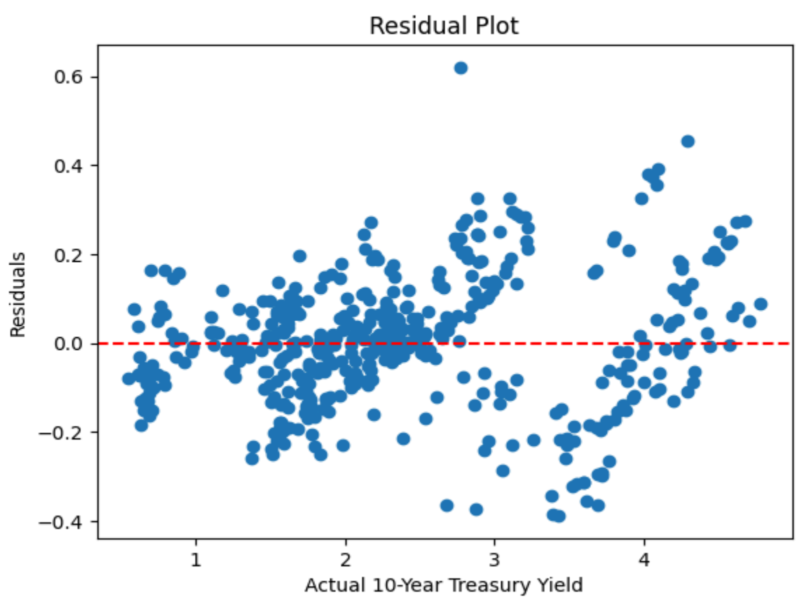
RMSE: 0.1423 R^2: 0.9801

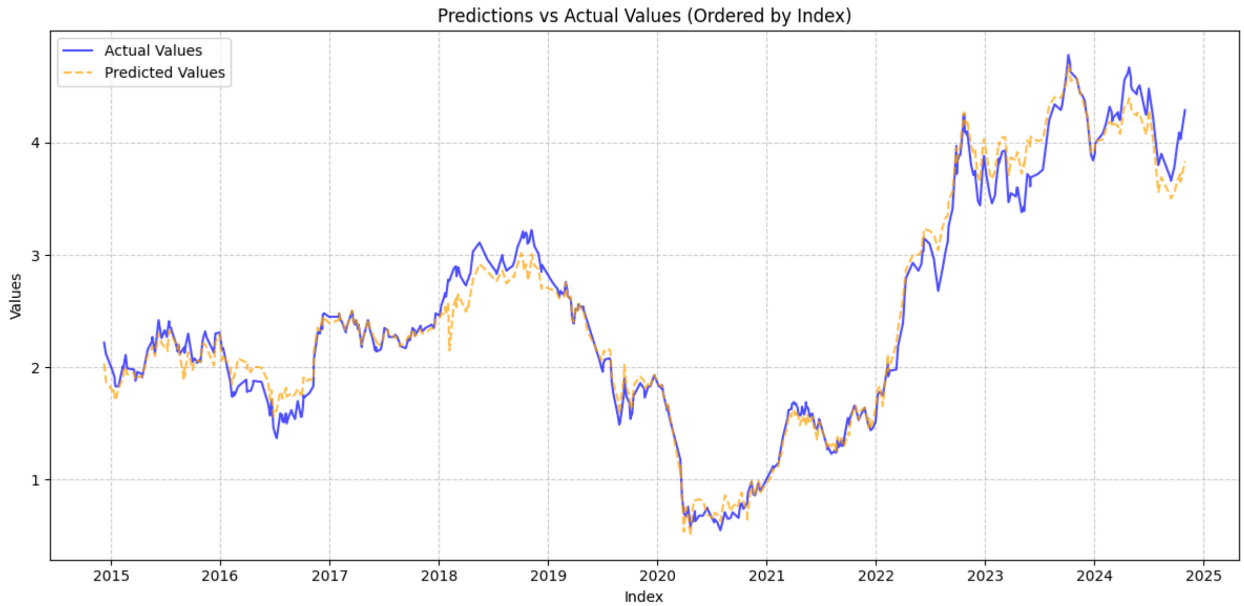
Folds Variability (Fold-Wise Metrics Plot):

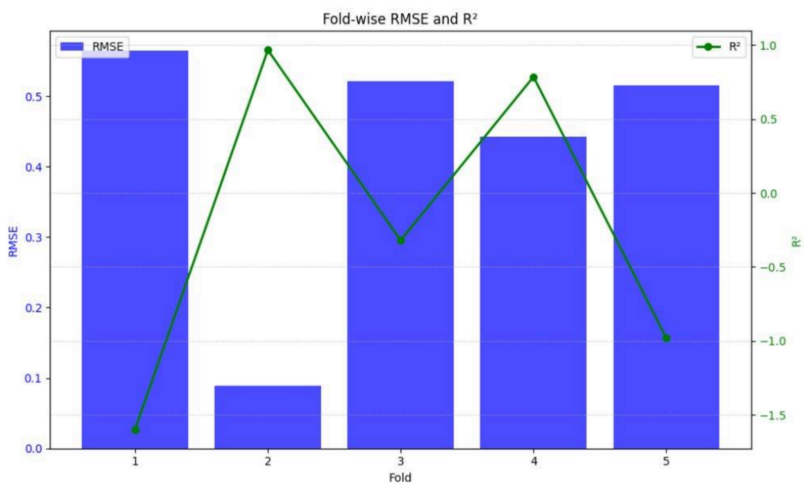
Mean RMSE: 0.4267 Mean R^2: -0.2297

The model performs very poorly because it struggles with most folds since the model is not complex enough to capture non-linear tendencies and temporal dependencies.


Actual vs Predicted Values
Prediction
Ideal Fit
Predicted Values
Actual Values
1
2
3
4


Residual Plot
Residuals
Actual 10-Year Treasury Yield
0.6
0.4
0.2
0.0
−0.2
−0.4
1
2
3
4





The regression analysis demonstrates the relationship between the 10-Year Treasury Yield and five independent macroeconomic indicators. The top scatter plot illustrates the predicted vs. actual values, showing a strong linear trend closely aligning with the ideal fit (red dashed line), indicating accurate model predictions. However, time series cross validation evaluates model performance and reflects the model's explanatory power. Most folds have highly negative R^2 which indicates that the model is not complex enough to capture non-linear tendencies and temporal dependencies. This analysis highlights room for improvement in consistency across data folds.

XGBoost

Due to the inconsistencies observed in the performance of the linear regression model across folds, particularly its limitations in capturing potential non-linear relationships between the 10-Year Treasury Yield and the macroeconomic indicators, XGBoost was chosen as an alternative. XGBoost is a robust ensemble learning technique based on decision trees that iteratively trains weak learners (trees) and combines their predictions to minimize errors. Its ability to handle non-linear relationships, manage large datasets efficiently, robustness to outliers and noise and leverages gradient boosting to optimize performance makes it a superior choice for improving prediction accuracy and addressing the variability seen in linear regression results.

Algorithm:

1. The data is split into training (80%) and testing (20%) sets using train_test_split.
2. A baseline XGBoost model is created with optimized (using grid search) parameters like n_estimators, learning_rate, max_depth and subsample.
3. The model is trained using the training data.
4. Predictions are made on the test data.
5. Time Series Cross-validation is used to evaluate the model's generalizability.
6. Feature importance is visualized using plot_importance.
7. Selected top 3 most important features using f-score.
8. Retrained the model (trying to achieve dimension reduction).

Parameter Tuning:

A grid search (GridSearchCV) is used to find the optimal hyperparameters:

n_estimators: Number of boosting rounds.

max_depth: Maximum depth of each tree.

learning_rate: Step size for weight updates.

subsample: Fraction of samples used for training each tree.

The best parameters are extracted and used to re-train the model.

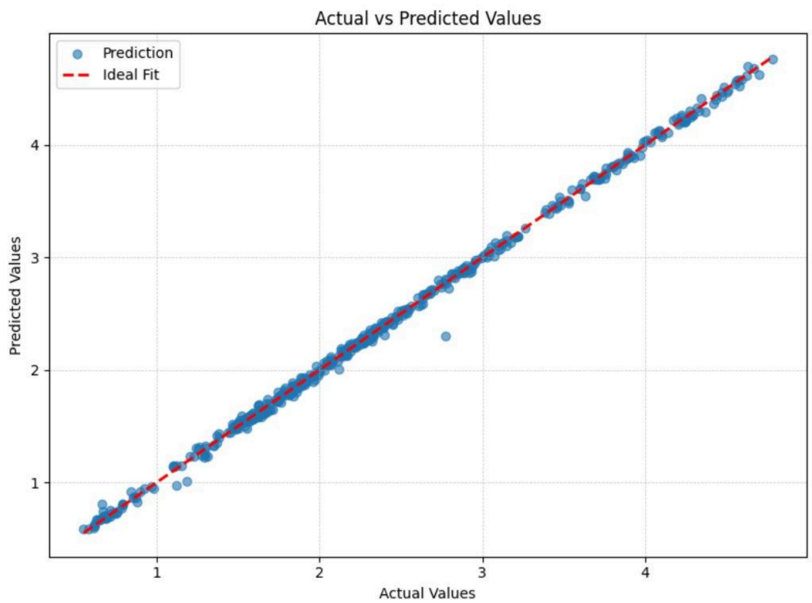
learning_rate: 0.1, max_depth: 5, n_estimators: 200, subsample: 0.7

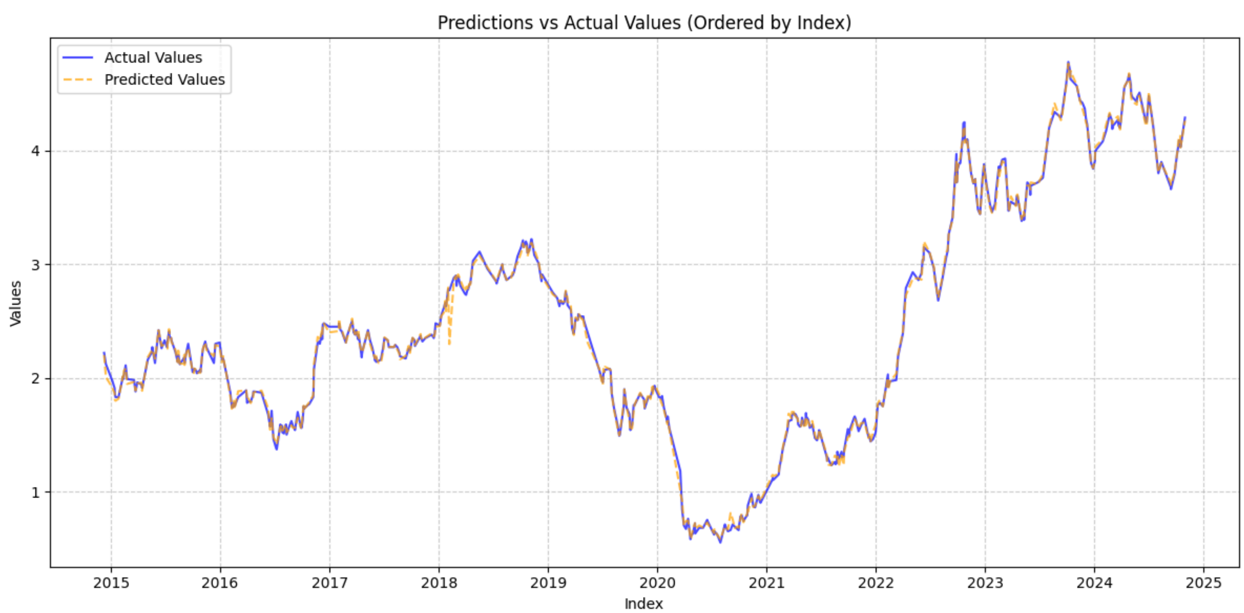
Model Performance:

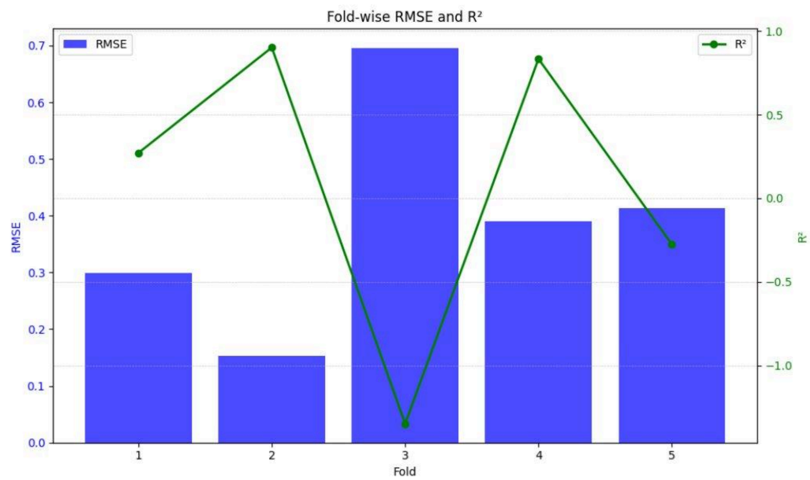
RMSE: 0.0387 R^2: 0.9985

Folds Variability (Fold-Wise Metrics Plot):

Mean RMSE: 0.3903   Mean R²: 0.0765







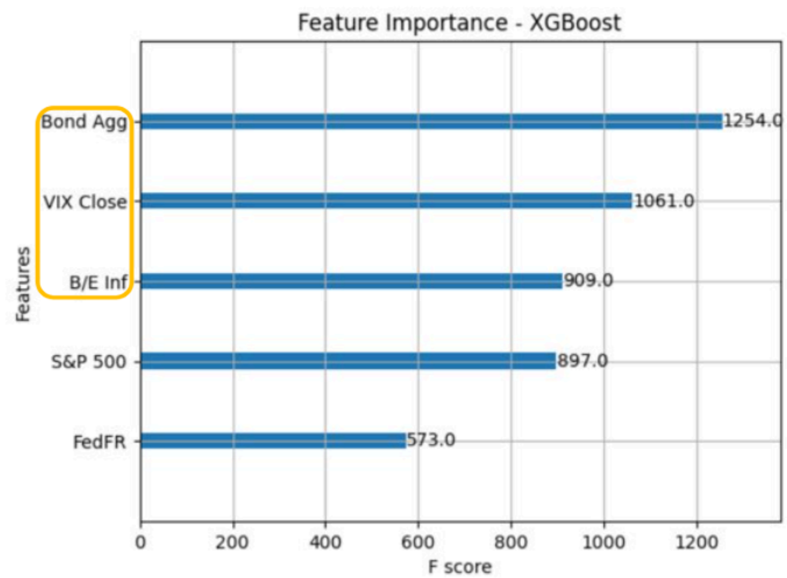
The XGBoost model results demonstrate a significant improvement in predictive accuracy and consistency compared to the linear regression model. The top scatter plot illustrates that the predictions align closely with the ideal fit (red dashed line), indicating that the model effectively captures the relationship between the dependent variable (10-Year Treasury Yield) and the independent macroeconomic indicators. This alignment highlights the model's overall robustness in predicting the target variable.

However, time series cross validation evaluates model performance and reflects the model's explanatory power. RMSE values remain relatively low across folds, signifying minimal prediction error. While there is some variability in R^2 across folds, the overall performance reflects the model's capability to handle non-linear and complex relationships among the five independent variables effectively, although it still does not completely capture temporal tendencies especially since it struggles to model the underlying complexities in the 3rd fold corresponding to the COVID financial crisis.

These results reaffirm the suitability of XGBoost for this context, where traditional linear methods may fail to capture non-linear interactions and dependencies within the data.

XGBoost Simplified

The F-score of a feature is the number of times the feature is used in a split across all trees in the XGBoost model. A higher F-score means the feature is used more frequently for splits, indicating it has a greater influence on model predictions. A lower F-score means the feature is less influential.



Here we retrain the model by selecting the top 3 most important features based on their f-scores: *Bond Aggregate, VIX Close, Breakeven Inflation*.

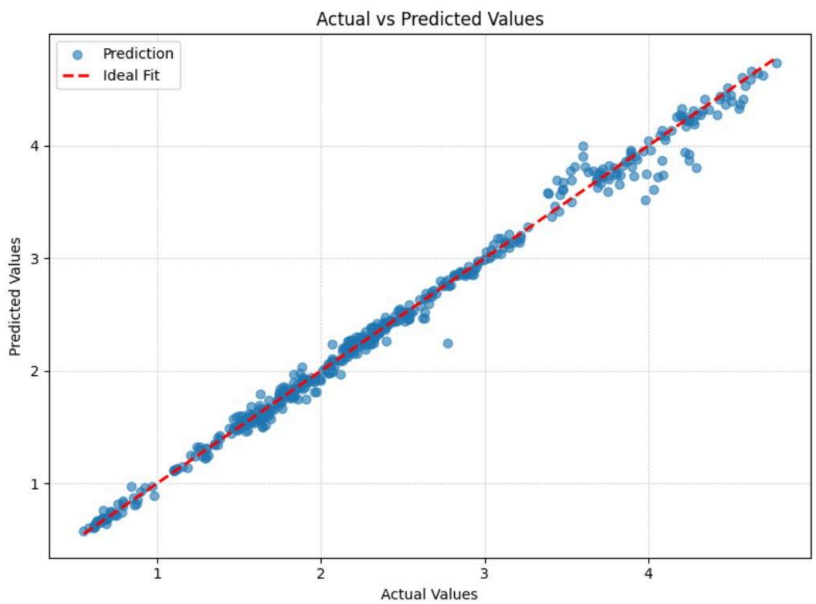
Model Performance:

The model achieves almost as good results on reducing number of features to 3, hence simplifying computation via dimension reduction.

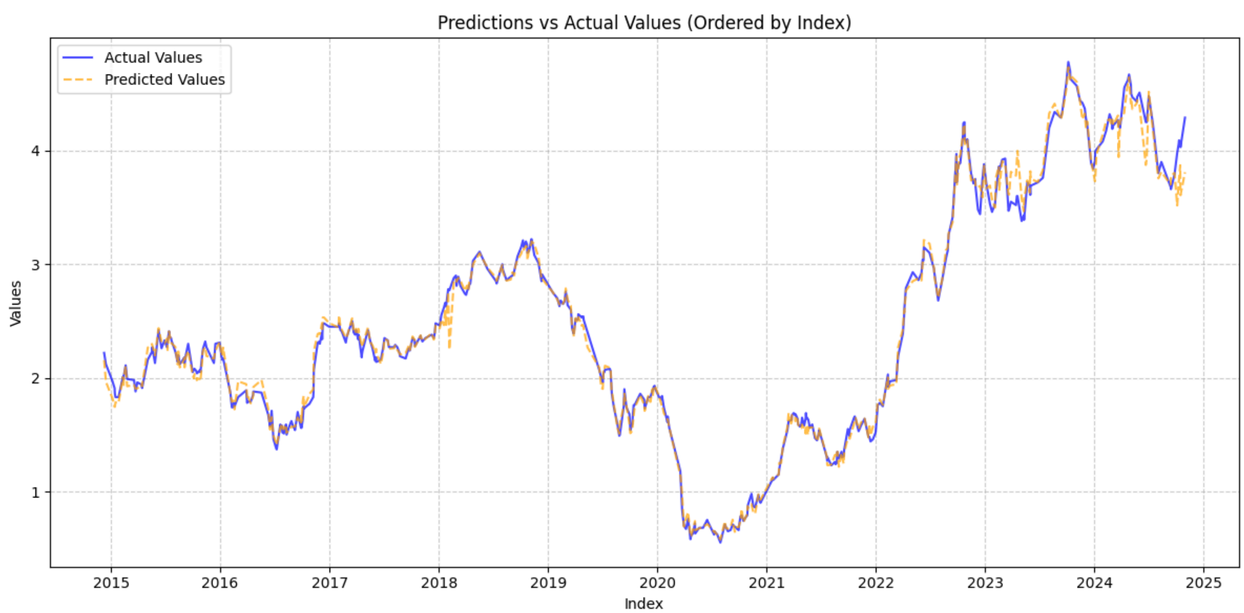
RMSE: 0.0862 R^2: 0.9927

Folds Variability (Fold-Wise Metrics Plot):

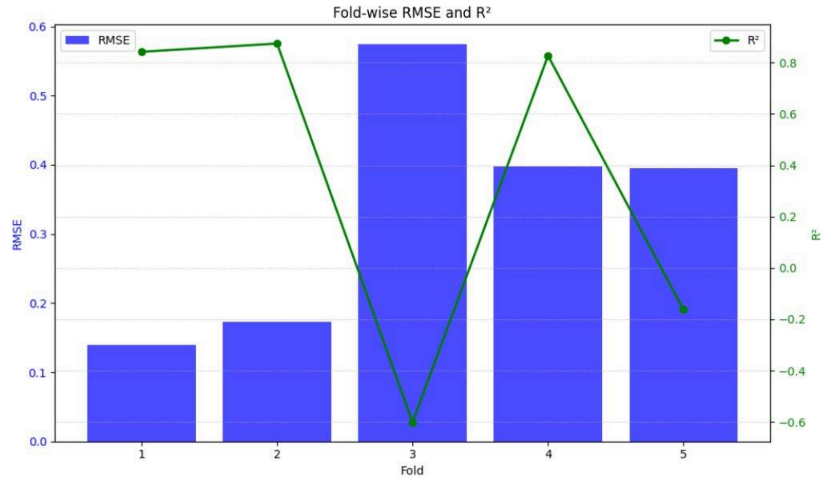
Mean RMSE: 0.3358   Mean R^2: 0.3559



Actual vs Predicted Values
Prediction
Ideal Fit
Predicted Values
Actual Values



Predictions vs Actual Values (Ordered by Index)
Actual Values
Predicted Values
Values
Index



The performance obtained by training the XGBoost model using only the top 3 most important features Bond Aggregate, VIX Close, Breakeven Inflation, as identified by feature importance scores in the previous model, demonstrates competitive results with reduced feature complexity.

The predictions remain closely aligned with the ideal fit line, indicating that the model successfully captures the underlying relationship even with fewer variables. This suggests that the top 3 features carry most of the predictive power for modeling the 10-Year Treasury Yield, and redundant or less important features do not significantly contribute to prediction accuracy.

However, time series cross validation evaluates model performance and reflects the model's explanatory power. RMSE values remain consistently low, implying that the model maintains its ability to minimize prediction errors. The R^2 values, mostly positive and relatively high, indicate that the model explains a substantial portion of the variance in the 10-Year Treasury Yield using just the most significant features. As observed previously, it does not completely capture temporal tendencies especially since it struggles to model the underlying complexities in the 3rd fold corresponding to the COVID financial crisis.

This result highlights the robustness of XGBoost in identifying and leveraging key drivers of the dependent variable. It also underscores the potential for feature selection to simplify models without sacrificing accuracy, which can be advantageous for interpretability and computational efficiency.

LSTM

Long Short-Term Memory (LSTM) networks, a specialized type of Recurrent Neural Network (RNN), are designed to model temporal dependencies in sequential data. Among the various models explored, including ARIMA and XGBoost, the LSTM yielded the best results in explaining the relationships between the yield and its driving variables, showcasing its ability to handle both temporal and multivariate data effectively.

Data Preprocessing

The first step in implementing the LSTM model involved preparing the data for sequential learning. All variables, including the dependent variable (10-Year Treasury Yield) and independent variables such as the Federal Funds Rate, VIX, and Breakeven Inflation Rate, were normalized using the MinMaxScaler to a range of [0, 1]. This normalization ensured numerical stability and faster convergence during training.

To enable the model to learn temporal patterns, a sliding window approach was employed. For each prediction, a sequence of the previous 60 days of data was used as input to predict the Treasury Yield for the next day. The 60-day sequence length was chosen after experimentation to strike a balance between capturing sufficient temporal context and avoiding overfitting. Each input sequence incorporated all independent variables, allowing the model to capture not only the temporal trends in the yield but also the interdependencies among the macroeconomic factors.

Model Architecture

The LSTM model's architecture was carefully designed to balance complexity and performance. It consisted of two stacked LSTM layers, each with 50 units, which processed the input sequences to extract temporal patterns. A Dense output layer was used to generate a single prediction for the next day's Treasury Yield. The model was compiled using the Adam optimizer, which is well-suited for efficient gradient descent, and the Mean Squared Error (MSE) loss function, ideal for continuous regression tasks.

Hyperparameter tuning played a critical role in optimizing the model's performance. Using Keras Tuner, key parameters such as the number of LSTM units, learning rate, sequence length, and batch size were systematically adjusted. The best-performing configuration included 50 LSTM units, a learning rate of 0.001, a sequence length of 60 days, and a batch size of 32.

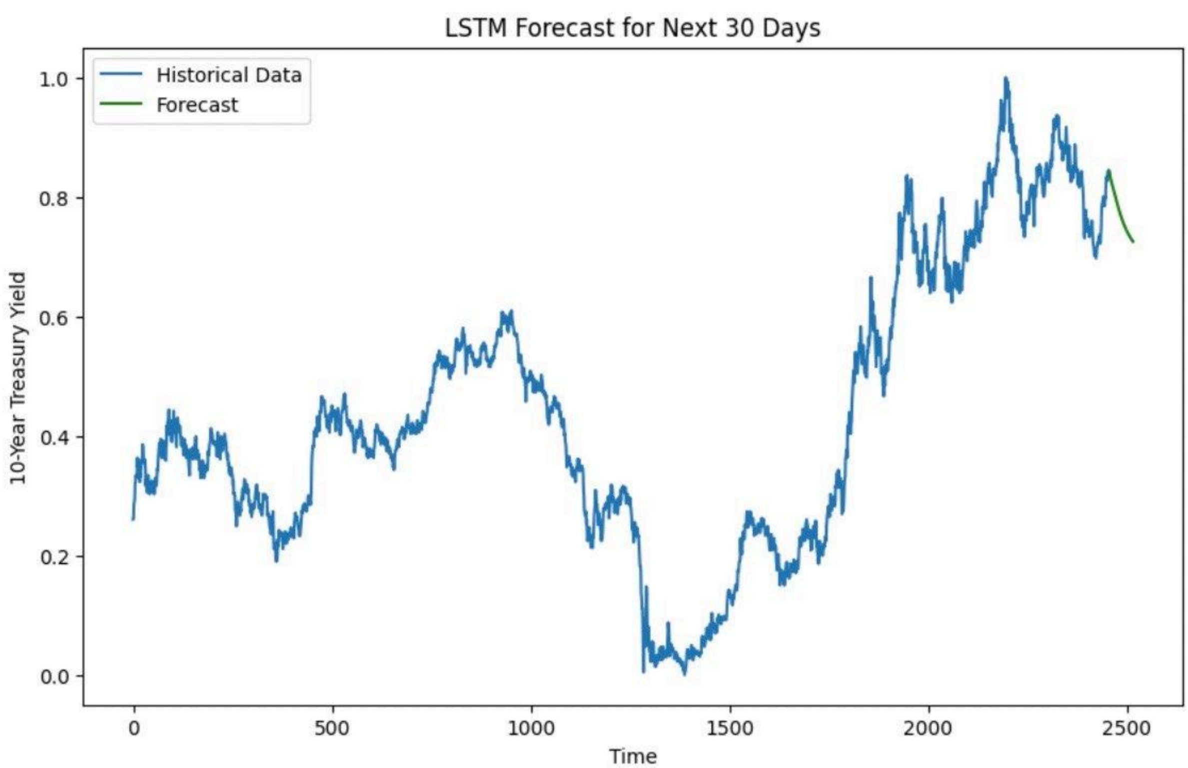
Training and Testing

The dataset was split into 80% training data and 20% testing data to evaluate the model's ability to generalize to unseen data. The model was trained over 50 epochs, with early stopping implemented to prevent overfitting by monitoring the validation loss. During training, the LSTM model processed the multivariate sequences to learn temporal and cross-variable dependencies, enabling it to explain the yield’s movements effectively.

Performance Evaluation

The LSTM model demonstrated superior performance compared to the other models analyzed in this project. It achieved a Root Mean Squared Error (RMSE) of 0.0192 and an R^2 of 0.9487, outperforming ARIMA and XGBoost in both accuracy and explanatory power. Residual analysis revealed minimal systematic errors, further validating the model's ability to replicate historical patterns.

To evaluate the model’s capacity to project trends, a 30-day forecast was generated. This was done iteratively, where the model predicted one day at a time, using the most recent prediction as input for the next day’s forecast. The forecast closely aligned with recent trends, reflecting the model’s understanding of temporal patterns. However, it is important to note that the forecast assumes stable market conditions and may struggle to account for unexpected economic shocks or structural changes.



This figure illustrates the LSTM model's 30-day forecast (green line) for the 10-Year Treasury Yield, overlayed on the historical yield data (blue line). The model captures the underlying trends and patterns of the yield, providing a forward-looking estimate based on temporal dependencies in the data.

Key Observations

The LSTM model effectively captured both temporal dependencies and multivariate relationships, making it the most suitable tool for explaining the 10-Year Treasury Yield. Variables like the Federal Funds Rate and Breakeven Inflation Rate emerged as significant contributors, reinforcing their economic relevance. While the model performed well in explaining yield movements, its reliance on stable trends underscores the importance of considering external factors and shocks in future work.

Conclusion

Among the models tested, the LSTM model provided the most accurate and insightful explanation of the 10-Year Treasury Yield, highlighting its robustness in handling sequential and multivariate data. The findings demonstrate the value of LSTMs for uncovering the complex dynamics between macroeconomic indicators and yield movements, offering a foundation for further exploration and practical applications in financial analysis. While the model's focus was on explanation rather than long-term prediction, it lays the groundwork for integrating advanced machine learning techniques into fixed-income research.

Model ranking

LSTM > XGBoost > Random Forest + SARIMAX > Random Forest > Linear Regression

Limitations

The model relies heavily on the assumption that historical patterns will persist, which may not hold true during economic shocks or regime changes, potentially limiting its robustness in highly volatile scenarios. Additionally, its focus on smaller segments of data post-breakpoints raises concerns about overfitting, as it may overly tailor predictions to specific periods rather than capturing broader trends. The model is also highly sensitive to hyperparameters, such as the number of units, learning rate, and sequence length, necessitating extensive tuning to optimize performance. While the model demonstrates strong accuracy for short-term predictions, its effectiveness diminishes over longer forecast horizons due to the accumulation of prediction errors, highlighting the need for enhancements in its long-term forecasting capabilities.

Further Exploration

Further exploration of the model can focus on enhancing its robustness and accuracy by incorporating additional features and advanced modeling techniques. One approach is to incorporate more features, such as global economic indicators like unemployment rates and GDP growth, as well as non-economic factors like political instability or news sentiment, which may provide a broader context and improve predictive power. Hybrid modeling can also be employed by combining LSTM (Long Short-Term Memory) networks with simpler models, allowing for a balance between accuracy and interpretability, while capturing complex sequential dependencies. Additionally, alternative models such as Transformers or Temporal Fusion Transformers can be explored to better capture the intricate temporal relationships and dependencies present in financial time series data. Scenario testing can help evaluate the model's robustness by simulating the impact of hypothetical events, such as rate hikes or inflation shocks, to understand how the model responds to extreme market conditions. Finally, restricting feature values with thresholds based on prior expectations and market perspectives can ensure the model remains grounded in realistic assumptions and market dynamics, improving its practical utility for decision-making.