Yutong Zhao
Professor Jacob Koehler
ECON-UB 232 2025 Fall: Data Bootcamp
12 December 2025

Forecasting Producer Price Index of U.S. Fuel Oil Using Classical and Hybrid Time Series Models

1. Introduction

This project develops predictive models for the monthly Producer Price Index (PPI) of U.S. fuel oil using both classical time series models and machine learning techniques. The objective is to improve medium-term inflation forecasting accuracy by combining traditional univariate approaches with exogenous macroeconomic drivers, including consumer prices, oil spot prices, and industrial production.

Multiple benchmark models were evaluated, including naive baselines, Seasonal ARIMA (SARIMA), Exponential Smoothing (ETS), an ensemble of classical models, and a two-stage hybrid model that applies XGBoost to correct ETS residual errors using external economic variables.

The results demonstrate that the hybrid ETS + XGBoost model consistently produced the lowest forecasting error (MAPE) among all tested approaches, outperforming classical univariate and ensemble models.

2. Data Description

- Target Variable：

Fuel Oil Producer Price Index (PPI): APU000072511

- Exogenous Variables

Unemployment Rate: UNRATE
Consumer Price Index: CPIAUCSL
Crude Oil Spot Price (WTI): DCOILWTICO
Industrial Production Index: INDPRO

These series were chosen based on economic theory linking energy price inflation to input costs, business activity, labor conditions, and consumer price pressures.

- Data Processing

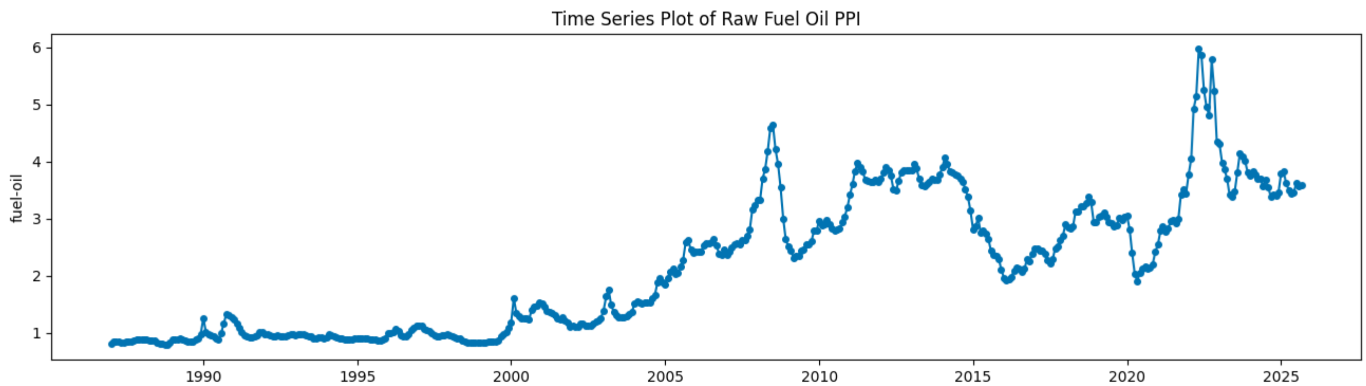
1) Missing values in exogenous data were resolved using linear interpolation and forward/backward filling. Below shows the information of the cleaned dataset:

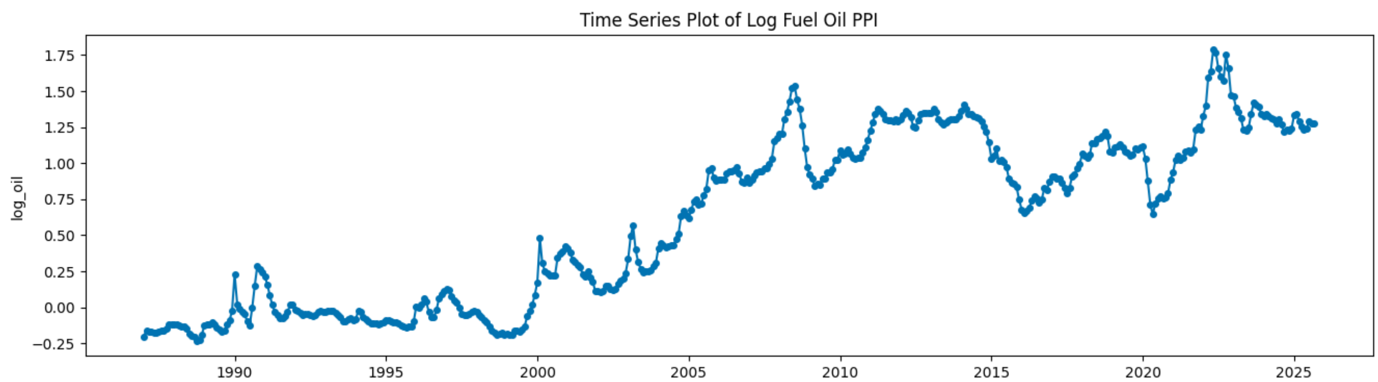
```
Data columns (total 5 columns):
#   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
0   fuel-oil                     465 non-null    float64
1   unemployment                 465 non-null    float64
2   cpi                          465 non-null    float64
3   Crude Oil (WTI) Price        465 non-null    float64
4   Industrial Production Index  465 non-null    float64
```

2) Low-correlation exogenous variables were screened out by computing their correlations with the target PPI series, and only variables with an absolute correlation coefficient greater than 0.4 were retained for modeling. The correlations are shown below:

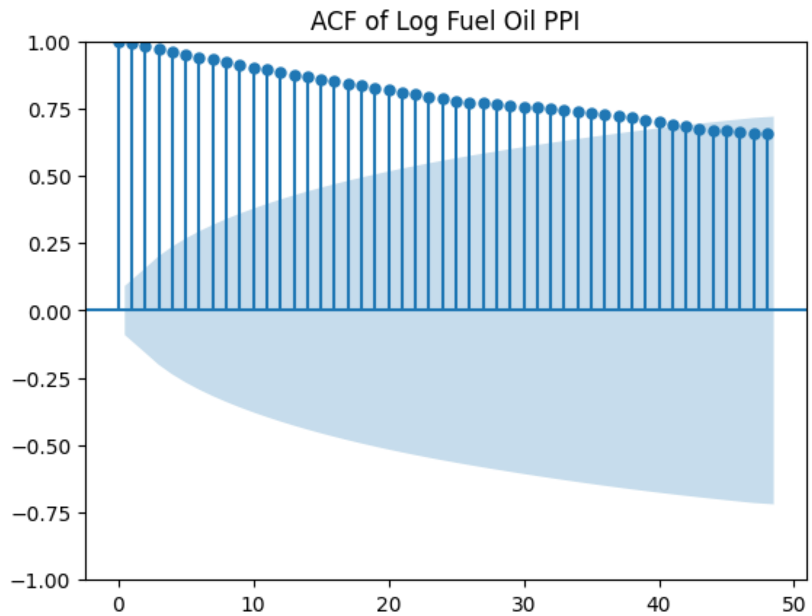
```
Correlation with PPI:
unemployment                  0.003260
cpi                           0.885557
Crude Oil (WTI) Price         0.938668
Industrial Production Index   0.826716
dtype: float64
Selected Exogenous Features: ['cpi', 'Crude Oil (WTI) Price',
'Industrial Production Index']
```

3) Log-transformation was used to remove level-dependent volatility in the original PPI time series.



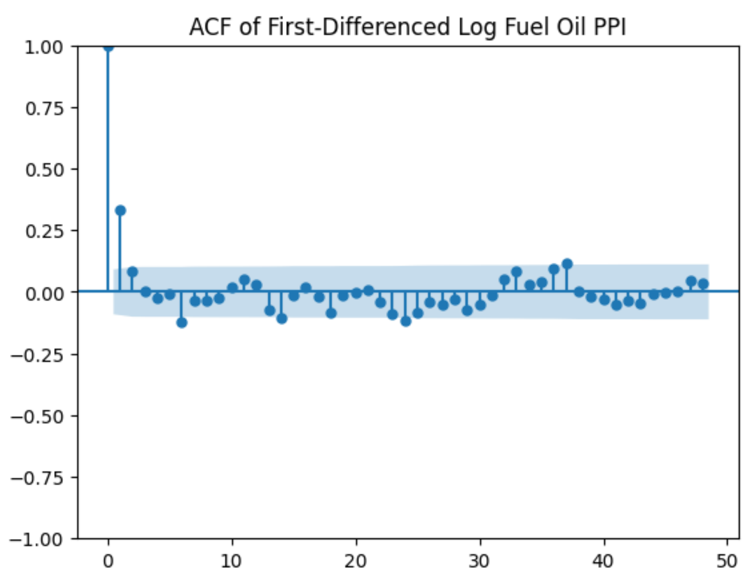


4) Stationarity was evaluated by plotting the ACF of Log PPI and also performing the Augmented Dickey-Fuller tests (ADF). The ACF plot shows significant autocorrelations, and the p-value in the ADF test is greater than 0.05, suggesting that the time series lacks stationarity. So the first difference was taken to achieve stationarity.




```
Original Series Stationarity Test:
ADF Statistic: -1.4299, p-value: 0.5679
```

The ACF plot and ADF test of Log PPI after differencing show stationarity:




```
Differenced Series Stationarity Tests:
ADF Statistic: -15.2333, p-value: 0.0000
```

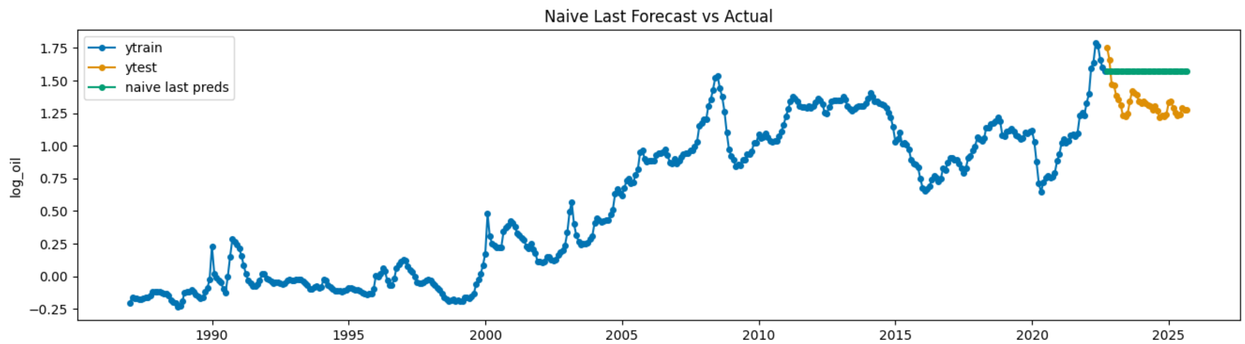
3. Models and Methods

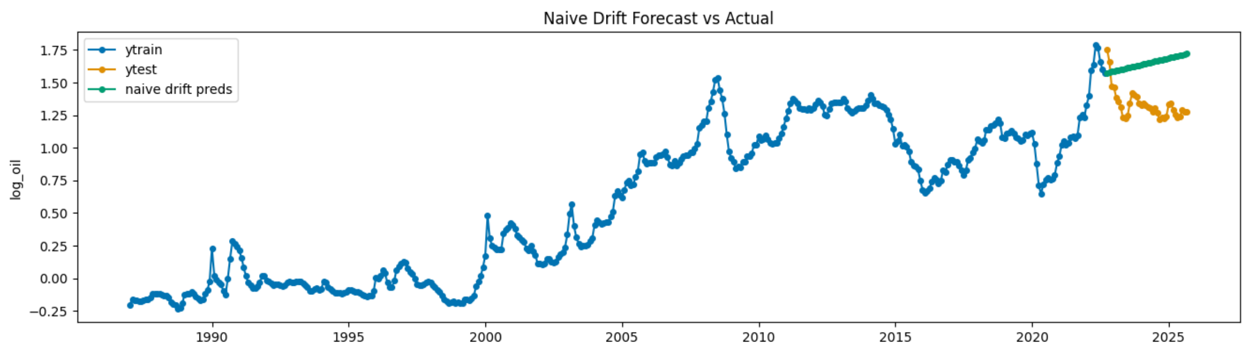
3.1 Benchmark Models

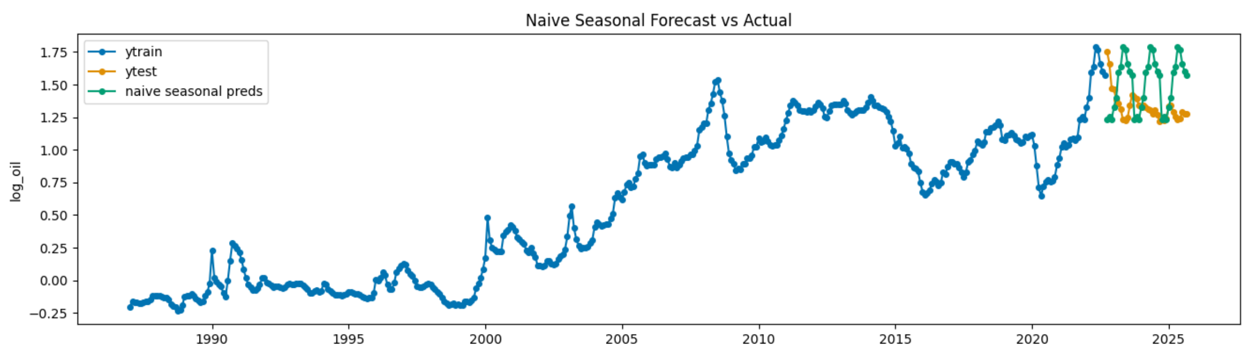
Several baseline probabilistic forecasting models were evaluated:

1) Naive Last Value
2) Naive Drift
3) Naive Seasonal (12-month repetition)

These serve as error floors for evaluating the effectiveness of more advanced techniques. The graphs of the results of the benchmark models are shown below:





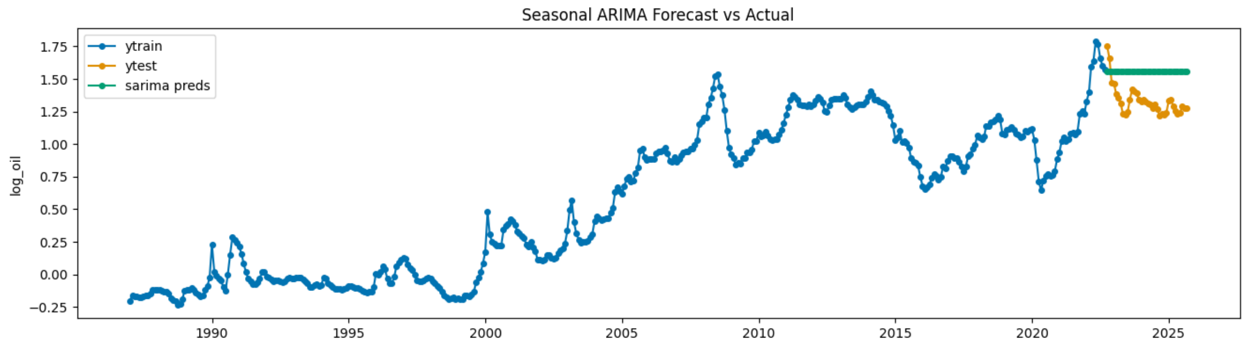


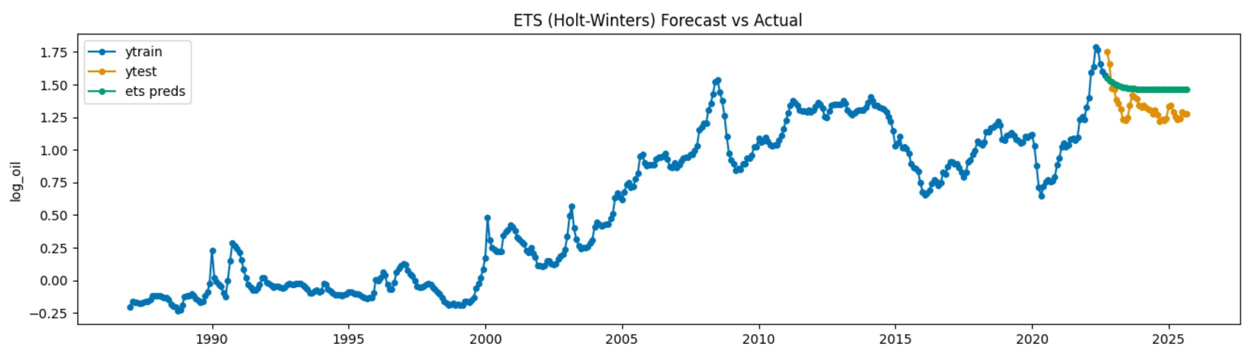
3.2 Classical Models

1) SARIMA: Seasonal ARIMA models were fitted using automated hyperparameter selection via AutoARIMA.
   - Seasonal period: 12 months
2) ETS: Automatic Holt-Winters exponential smoothing using AutoETS
   - Searches over additive/multiplicative trend and seasonality combinations
   - Seasonal length: 12

The graphs of the results of our classical models are shown below:

```
Best ARIMA model:  ARIMA(1,1,0)(0,0,0)[12]
```



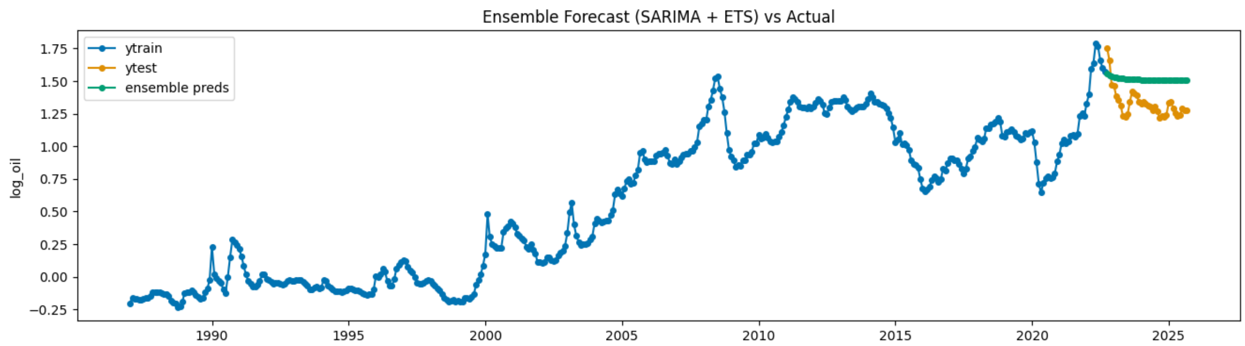


3.3 Ensemble Model

An ensemble model was built to combine SARIMA and ETS forecasts:

$$\hat{y}_t^{ensemble} = \frac{\hat{y}_t^{SARIMA} + \hat{y}_t^{ETS}}{2}$$

This approach aimed to average structural model uncertainty and reduce single-model bias. The graph of the results of the ensemble model is shown below:



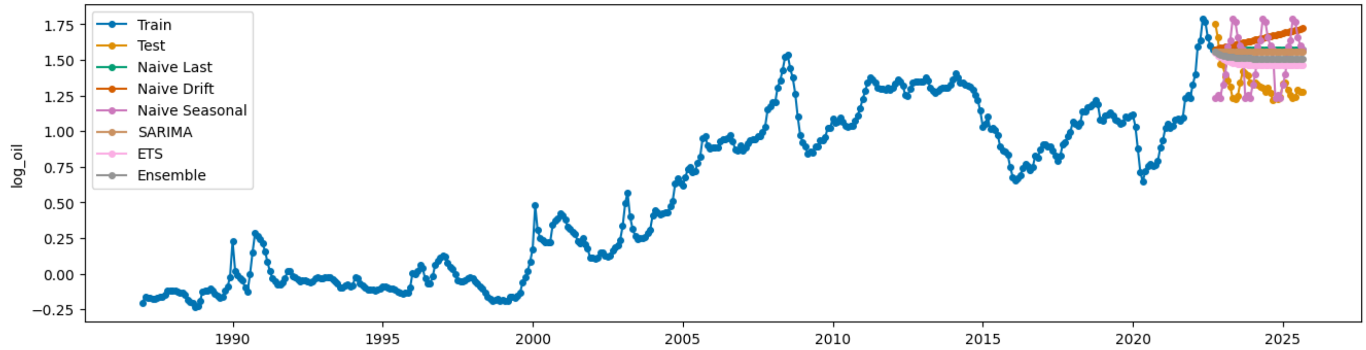
3.4 Hybrid ETS + XGBoost Model

To fully utilize macroeconomic data while preserving time-series structure, a two-stage hybrid forecasting strategy was implemented:

Step 1 — Baseline Forecast

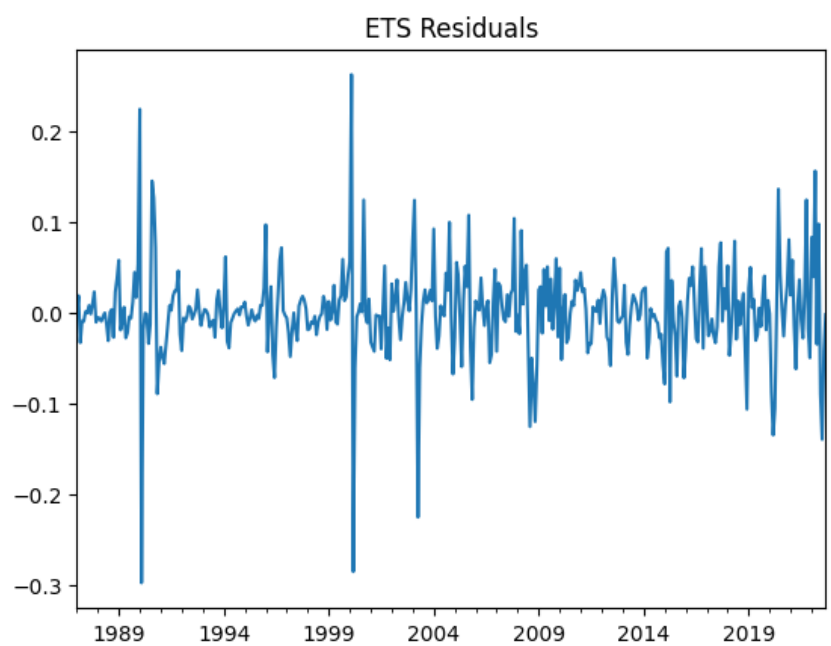
The models above are using only the historical PPI series. These models captured the core components of the time series, including trend, level, and seasonality, and produced baseline forecasts of PPI values. The model with the lowest mean absolute percentage error will be chosen for this step. As the results below, the ETS model has the lowest mean absolute percentage error (MAPE), so it is chosen as the baseline model for this step.

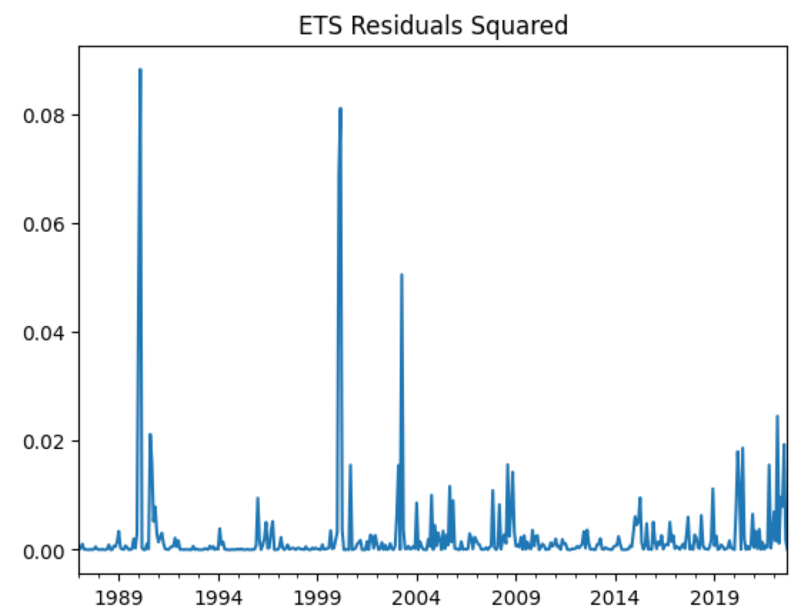
	Model	MAPE
0	Naive Last	0.197224
1	Naive Drift	0.256290
2	Naive Seasonal	0.205604
3	SARIMA	0.186237
4	ETS	0.124905
5	Ensemble	0.155571

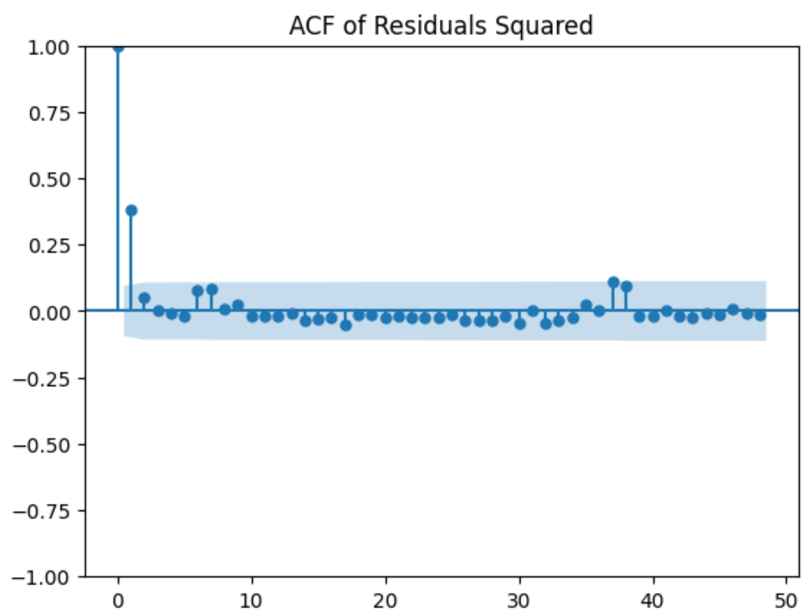


Step 2 — Residual Forecast

After generating ETS forecasts, the remaining errors (residuals) between the actual observations and the ETS predictions were computed. These residuals represent patterns not fully explained by classical smoothing methods, including nonlinear effects and external economic influences. After plotting the time series plot of residuals, it is clear that there are some significant spikes and clusters in the residuals. The time series plot and the ACF of squared residuals show autocorrelations in the residuals, so they are not pure white noise, which means they may be forecastable.





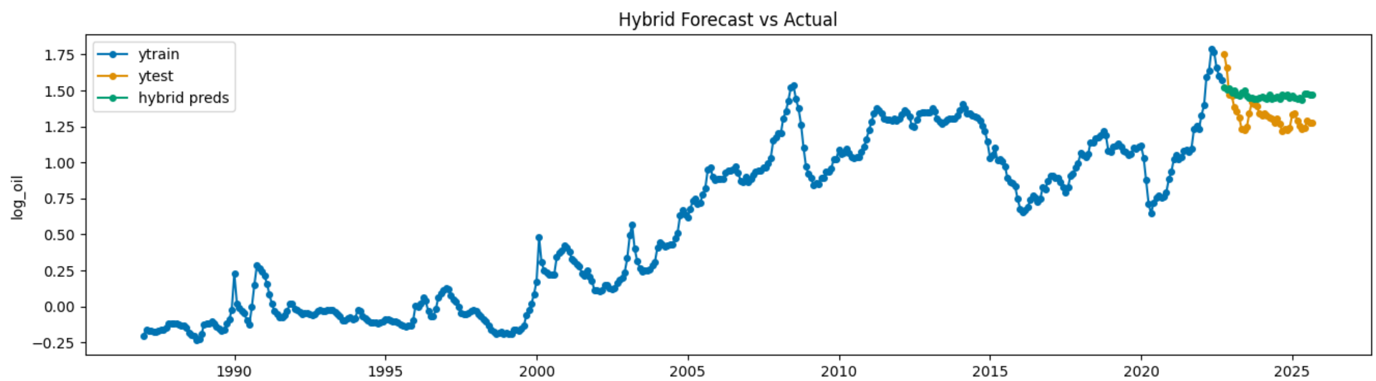


An XGBoost regression model was then trained to predict these residuals using scaled macroeconomic features, such as oil prices, consumer inflation, and industrial production. The model was configured with 300 decision trees, moderate depth to avoid overfitting, and subsampling strategies to improve generalization performance.

The trained XGBoost model produced forecasts of the estimated residual corrections for the forecast horizon.

Step 3 - Final Hybrid Forecast

The final hybrid forecast was obtained by adjusting the ETS baseline predictions using the residual corrections generated by the XGBoost model. This hybrid structure preserves the strength of classical exponential smoothing for modeling seasonality and trend while incorporating flexible machine-learning corrections driven by macroeconomic signals. This combined approach enables improved forecast accuracy by accounting for both intrinsic time-series components and complex, nonlinear exogenous influences that would otherwise remain unmodeled.



4. Results and Interpretation

4.1 Evaluation Methodology

Model performance was evaluated using a hold-out temporal test set consisting of the final 36 months of observations. All models were trained exclusively on historical data prior to this period to preserve a realistic forecasting scenario.

Forecast accuracy was measured using Mean Absolute Percentage Error (MAPE). MAPE quantifies the average absolute deviation between predictions and actual observations as a percentage of the true values, thereby providing an intuitive measure of relative error across competing models. Lower MAPE values indicate superior performance.

4.2 Results Summary

Model performance ranked by MAPE from best to worst is shown below:

	Model	MAPE
6	Hybrid ETS + XGBoost	0.119802
4	ETS	0.124905
5	Ensemble	0.155571
3	SARIMA	0.186237
0	Naive Last	0.197224
2	Naive Seasonal	0.205604
1	Naive Drift	0.256290

The results show that combining the classical ETS model with XGBoost generates the lowest MAPE and outperforms models that only make predictions using a single target series.

4.3 Interpretation of Results

1) Hybrid ETS + XGBoost

The hybrid model achieved the lowest forecasting error, outperforming all baselines. ETS captured the trend and seasonal patterns in PPI, while XGBoost modeled the nonlinear effects of macroeconomic factors on residuals. This approach effectively captured complex dynamics—such as supply shocks and energy market cycles—while maintaining interpretability and improving accuracy beyond classical methods.

2) ETS

ETS alone performed strongly, reflecting the stable trend and seasonal structure of PPI. The modest gain from the hybrid model indicates that classical methods explain most variation, but external macroeconomic effects add predictive value.

3) Ensemble Model

The ensemble of ETS and SARIMA improved over SARIMA individually but did not outperform ETS on its own. This suggests that simple averaging only provides benefits when component models contribute complementary information; since ETS already outperformed SARIMA and their errors are correlated, averaging diluted the stronger ETS signal rather than enhancing it.

3) SARIMA

SARIMA achieved moderate accuracy. While it captures autocorrelation and seasonality, its linear assumptions and lack of exogenous variables limit its responsiveness to abrupt or nonlinear macroeconomic changes.

4) Naive Models

Naive forecasts performed poorly, confirming the presence of meaningful structure and external drivers beyond simple persistence or trends.

5. Conclusion and Next Steps

In conclusion, this study demonstrates the effectiveness of classical time-series models for seasonal economic forecasting. While models such as ETS and SARIMA capture trend, seasonality, and autocorrelation, incorporating macroeconomic drivers through machine learning provides improvements in predictive accuracy. The two-stage hybrid design, which combines ETS baseline forecasts with XGBoost residual corrections, outperforms direct approaches by capturing both intrinsic time-series structure and nonlinear effects from external economic variables.

Moving forward, model performance could be further enhanced by expanding the set of macroeconomic features, exploring alternative machine learning techniques for residual forecasting such as long short-term memory (LSTM) networks or other recurrent neural architectures, optimizing ensemble strategies, and incorporating probabilistic forecasts with end-to-end deep learning models, like pretrained Time Series Transformers or the Temporal Fusion Transformer in PyTorch, which would provide point forecasts, uncertainty estimates, and deeper insights into variable importance and nonlinear macroeconomic interactions. These steps would provide a more robust framework for medium - to long-term energy price inflation forecasting and support informed decision-making in energy markets.