Forecasting PPI for U.S. Fuel Oil Using Classical and Hybrid Time Series Models

Data Bootcamp Final Project

Yutong Zhao

Forecasting PPI for U.S. Fuel Oil Using Classical and Hybrid Time Series Models

Data Bootcamp Final Project

Yutong Zhao

Outlines

01
Introduction

02
Data Description

03
Models and Methods

04
Results and Interpretation

05
Conclusion and Next Steps

01

Introduction

Predictive Task:
Producer Price Index of U.S. Fuel Oil

Objective:
Compare different predictive models and choose the most accurate one for our prediction

02

Data Description

Data Sources and Variables

Target

- Fuel Oil Producer Price Index (PPI)

Exogenous Variables

- CPI
- Crude Oil WTI spot price
- Industrial Production Index
- Unemployment Rate

Data Processing

01
Data cleaning

Missing values in exogenous data were resolved using linear interpolation and forward/backward filling

02
Feature selection

We choose exogenous variables that have an absolute value of correlation with PPI greater than 0.4

03
Log transformation

To remove level-dependent volatility in the original PPI time series

04
Stationarity adjustment

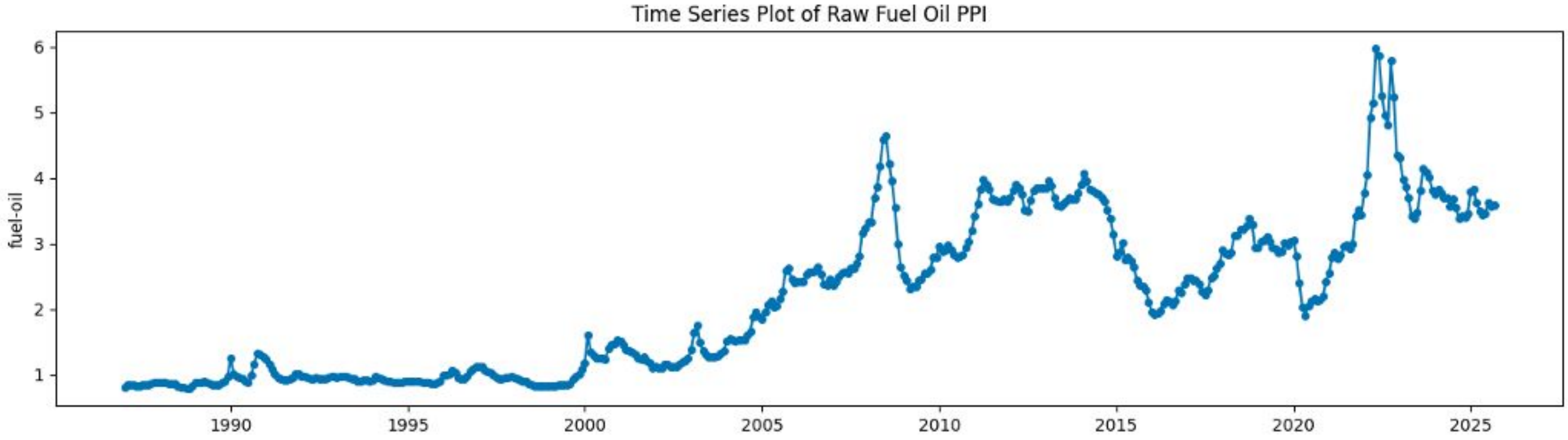
Take the first difference to achieve stationarity

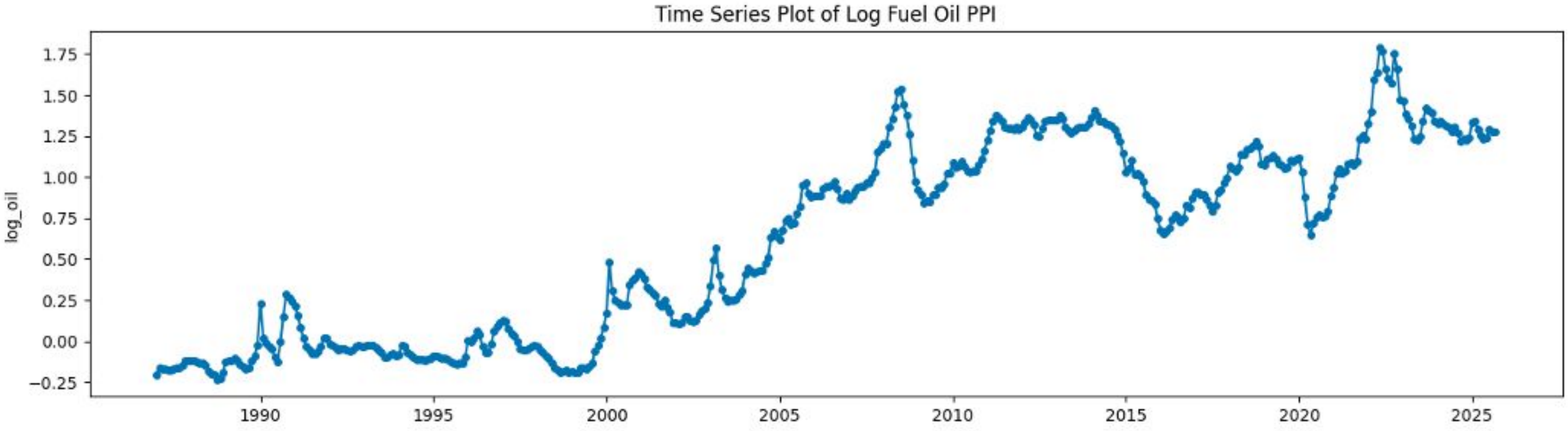
Feature Selection

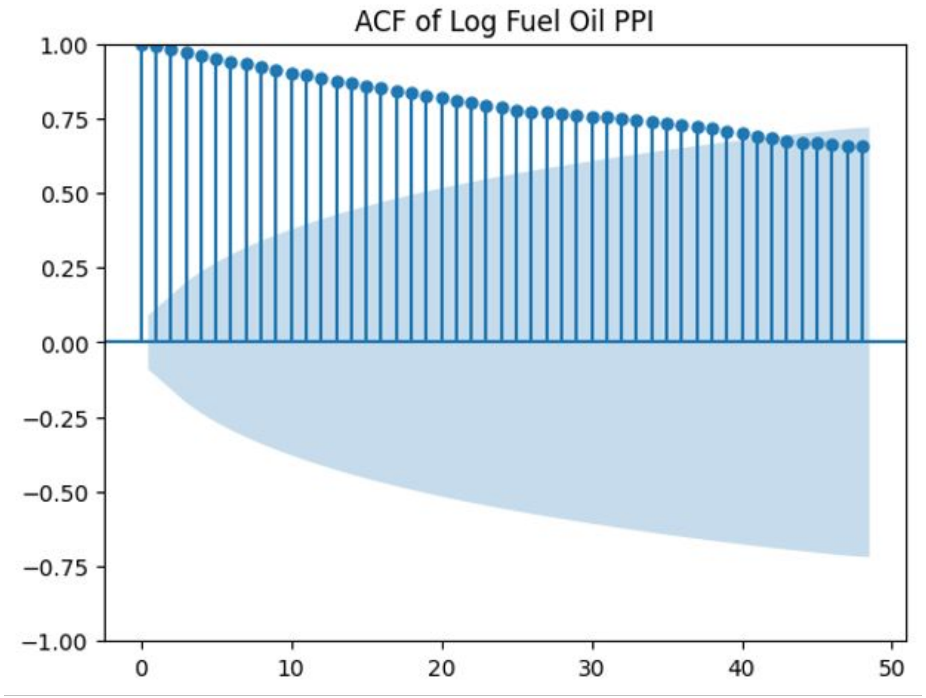
Correlation with PPI:

unemployment	0.003260
CPI	0.885557
Crude Oil (WTI) Price	0.938668
Industrial Production Index	0.826716

Final selected exogenous features: CPI, Crude Oil (WTI) Price, Industrial Production Index

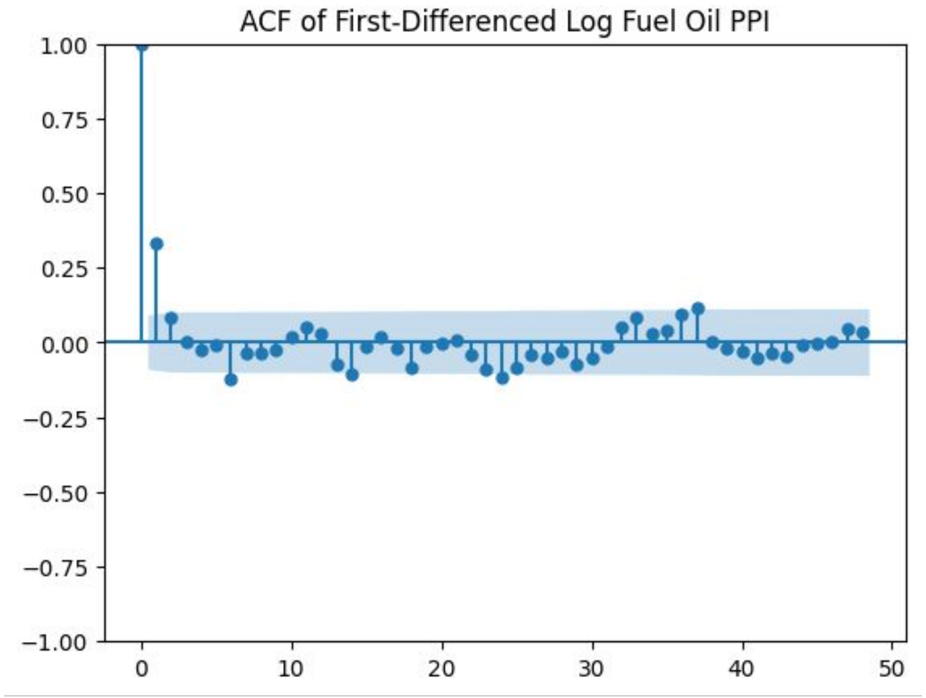

Time Series Plot of Raw Fuel Oil PPI
fuel-oil
1
2
3
4
5
6
1990
1995
2000
2005
2010
2015
2020
2025


Time Series Plot of Log Fuel Oil PPI
log_oil
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
1990
1995
2000
2005
2010
2015
2020
2025



Original Series Stationarity Test:

ADF Statistic: -1.4299, p-value: 0.5679



Differenced Series Stationarity Tests:

ADF Statistic: -15.2333, p-value: 0.0000

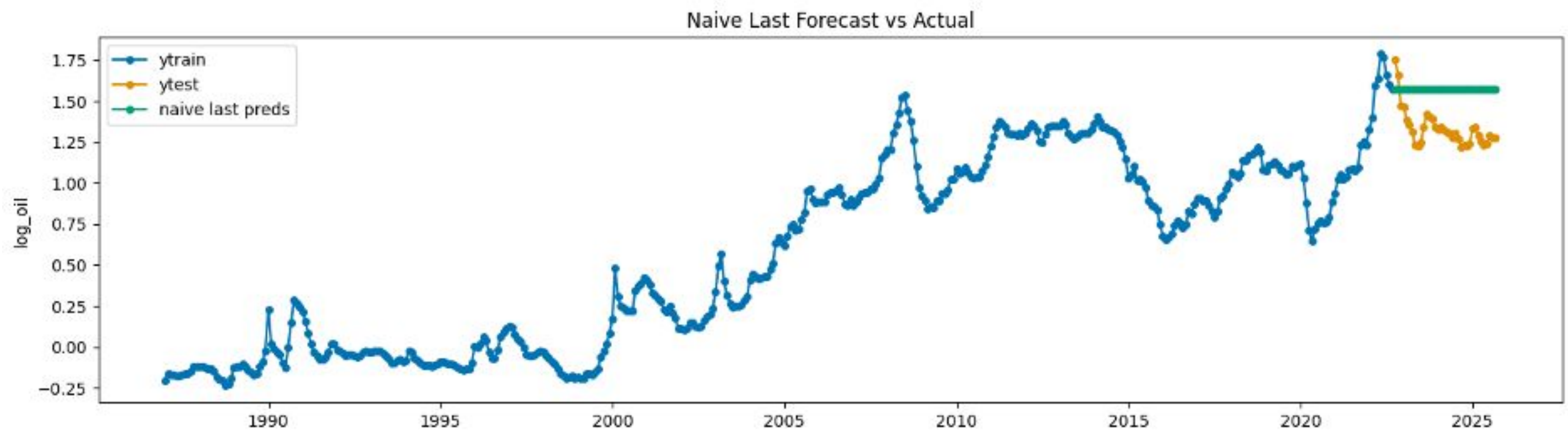
03

Models and Methods

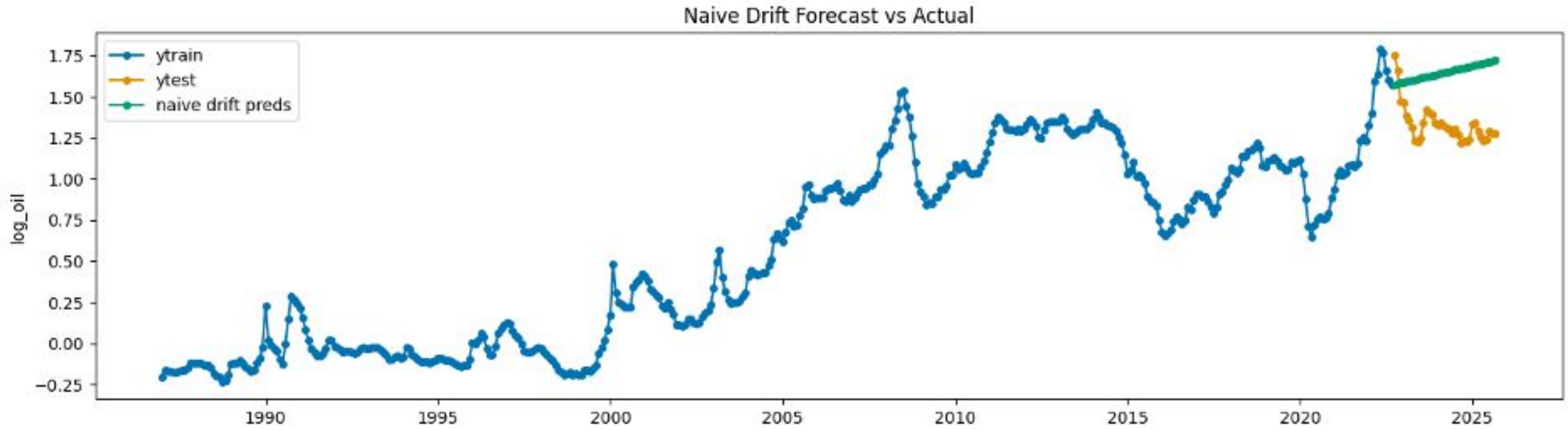
Benchmark Models

- Naive Last Value
- Naive Drift
- Naive Seasonal

Naive Last Value

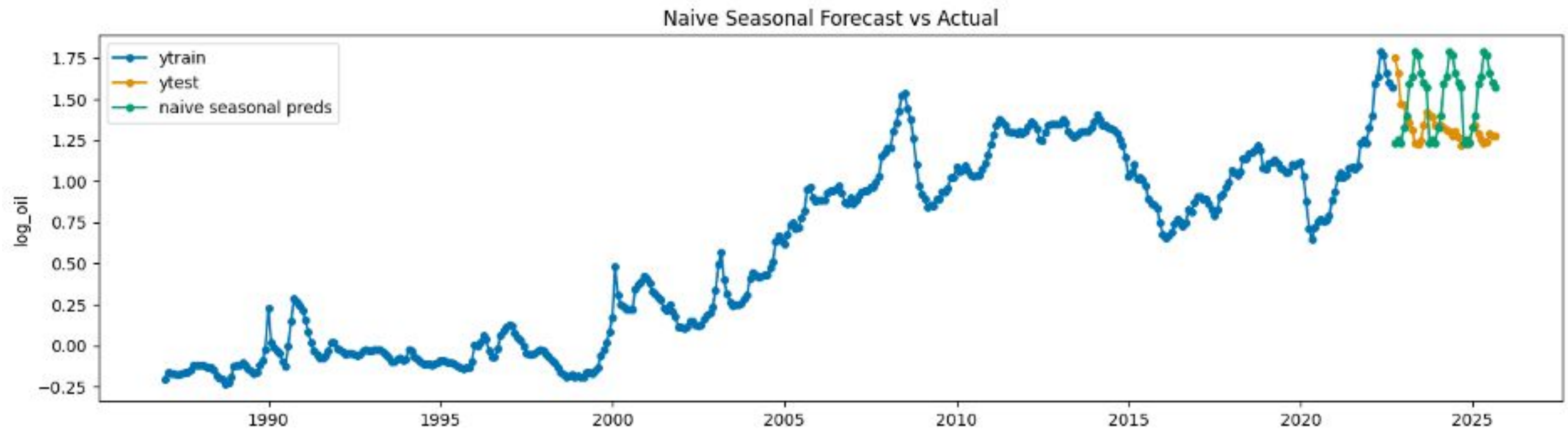


Naive Drift



Naive Drift Forecast vs Actual
ytrain
ytest
naive drift preds
log_oil
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
1990
1995
2000
2005
2010
2015
2020
2025

Naive Seasonal

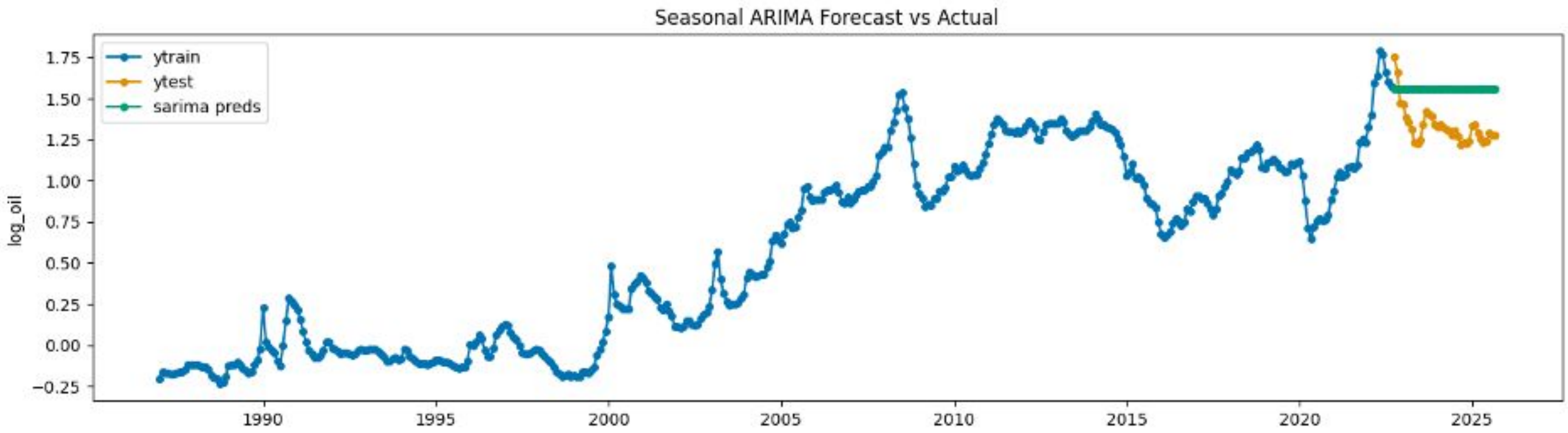


Classical Models

- **SARIMA**: Seasonal ARIMA models were fit using automated hyperparameter selection via AutoARIMA.

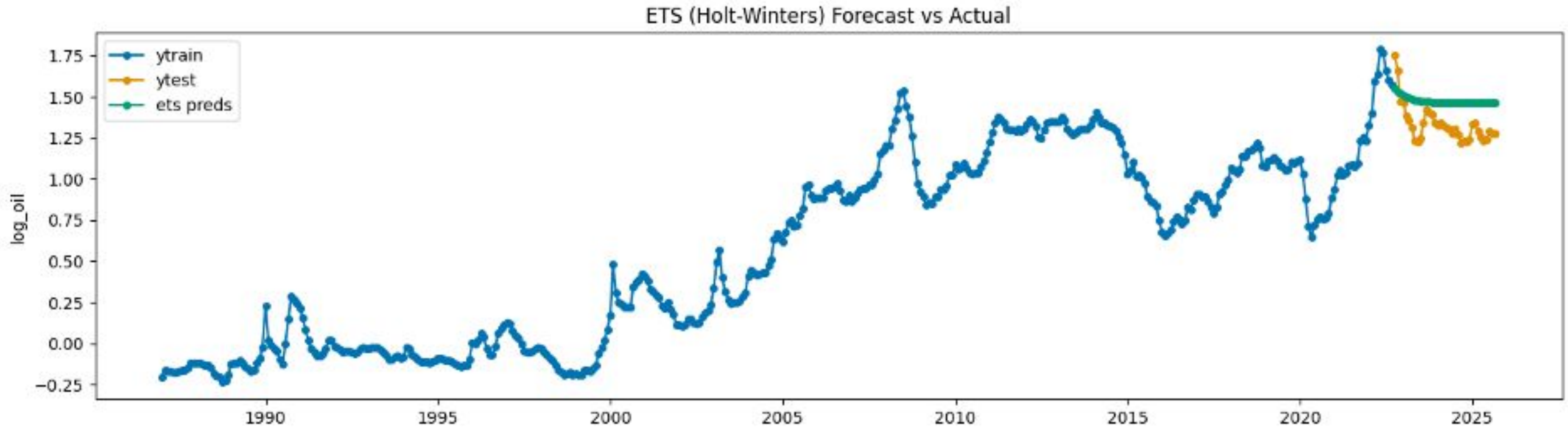
- **ETS**: Automatic Holt-Winters exponential smoothing using AutoETS

SARIMA

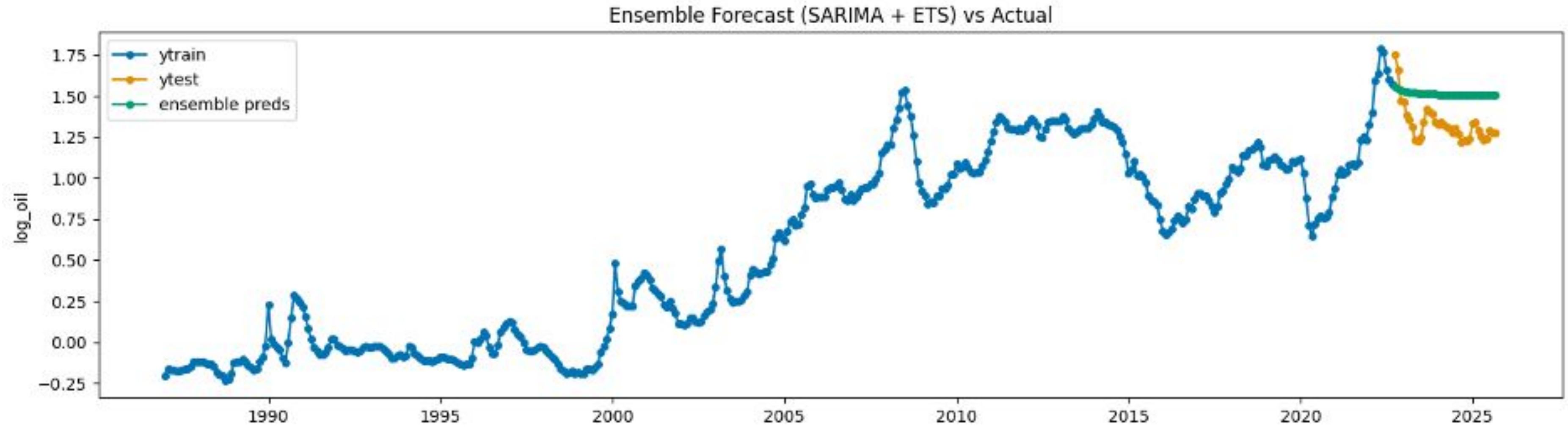


Seasonal ARIMA Forecast vs Actual
ytrain
ytest
sarima preds
log_oil
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
1990
1995
2000
2005
2010
2015
2020
2025

Exponential Smoothing (ETS)



Ensemble Model



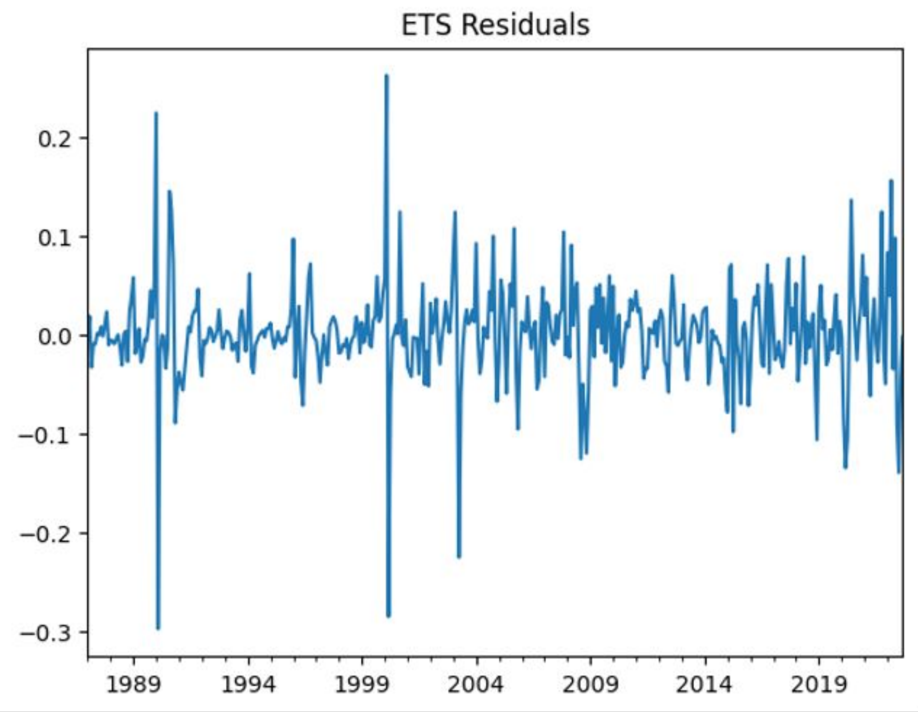
Hybrid ETS + XGBoost Model

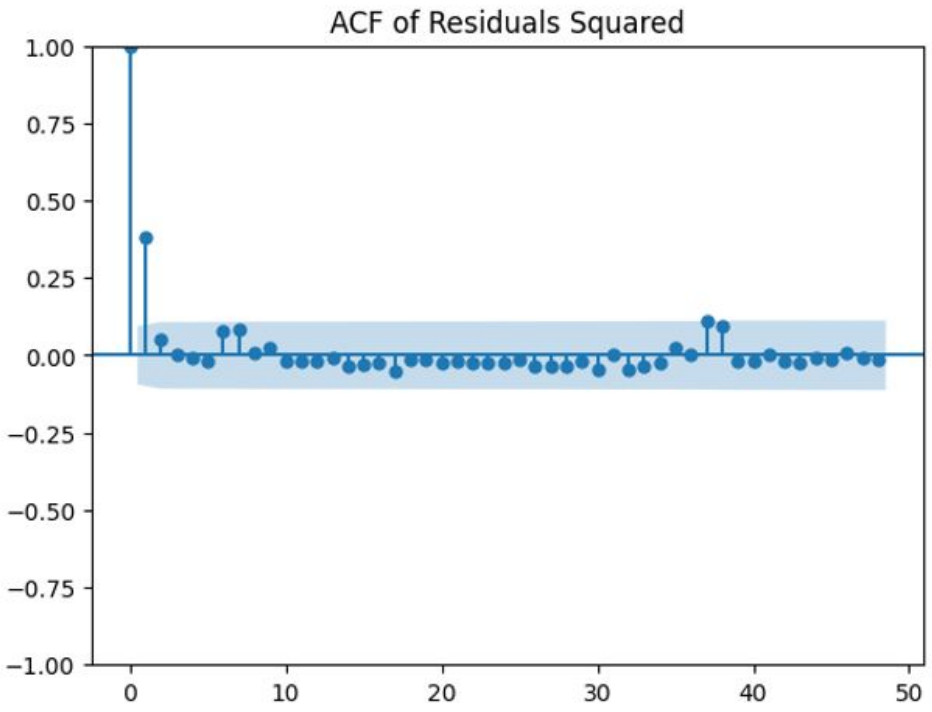
Step 1 – Baseline Forecast Using ETS Model

	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

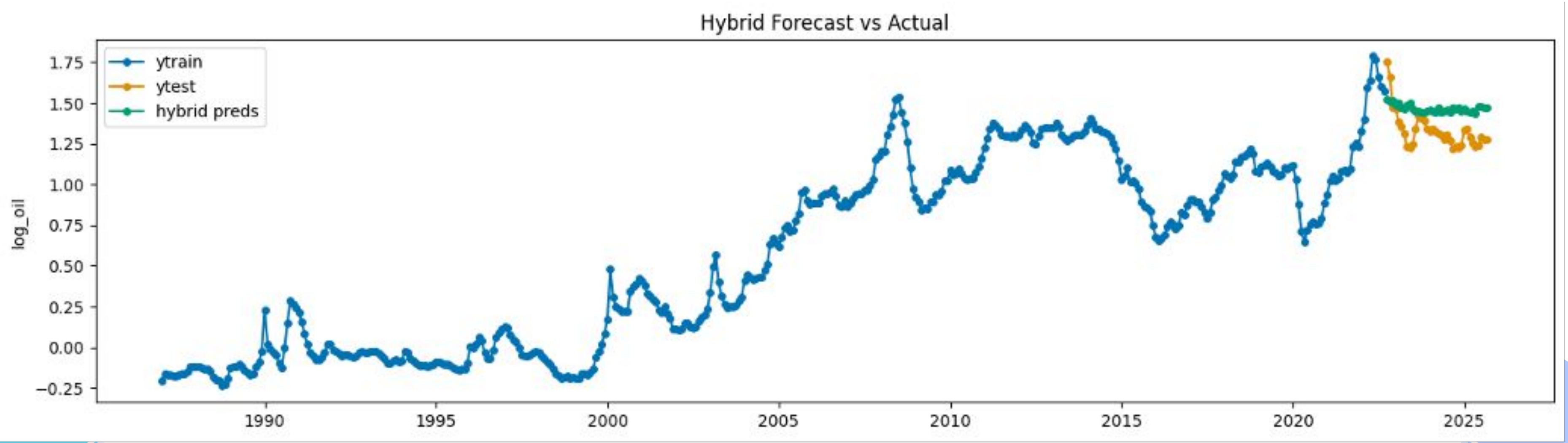
- Forecastability checking
- XGBoost Model


ETS Residuals
0.2
0.1
0.0
−0.1
−0.2
−0.3
1989
1994
1999
2004
2009
2014
2019


ACF of Residuals Squared
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50

Step 3 - Final Hybrid Forecast

ETS PPI forecast + XGBoost residual forecast



04

Results and Interpretation

Model Evaluation Method:
Mean Absolute Percentage Error (MAPE)

	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

Conclusion

- Classical time-series models such as ETS and ARIMA for seasonal economic forecasting are effective
- Incorporating macroeconomic drivers through machine learning provides improvements in predictive accuracy

Further Analysis

- Expand the set of macroeconomic features
- Explore alternative machine learning techniques for residual forecasting, such as neural networks
- Optimize ensemble strategies
- Incorporate probabilistic forecasts to quantify uncertainty

Thank you!

Outlines

01
Introduction

02
Data Description

03
Models and Methods

04
Results and Interpretation

05
Conclusion and Next Steps

01

Introduction

Predictive Task:
Producer Price Index of U.S. Fuel Oil

Objective:
Compare different predictive models and choose the most accurate one for our prediction

02

Data Description

Data Sources and Variables

Target

- Fuel Oil Producer Price Index (PPI)

Exogenous Variables

- CPI
- Crude Oil WTI spot price
- Industrial Production Index
- Unemployment Rate

Data Processing

01
Data cleaning

Missing values in exogenous data were resolved using linear interpolation and forward/backward filling

02
Feature selection

We choose exogenous variables that have an absolute value of correlation with PPI greater than 0.4

03
Log transformation

To remove level-dependent volatility in the original PPI time series

04
Stationarity adjustment

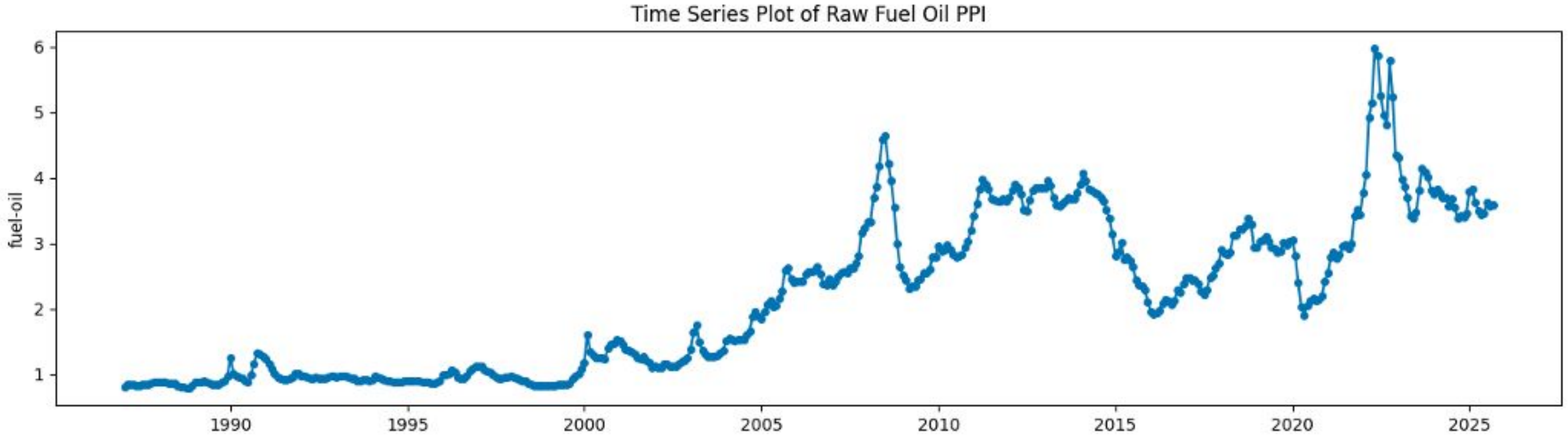
Take the first difference to achieve stationarity

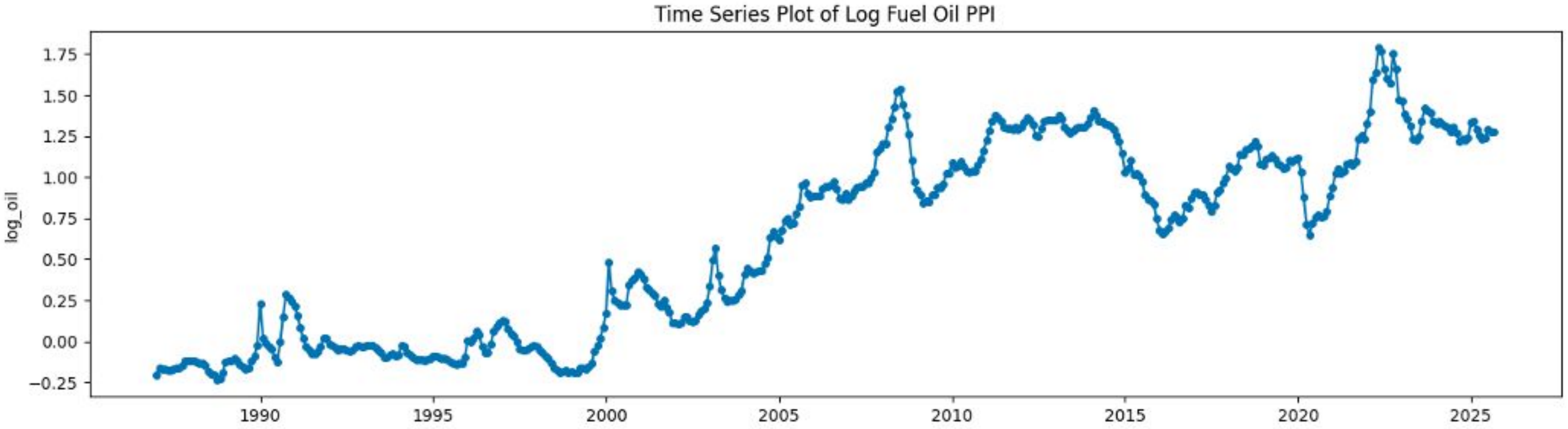
Feature Selection

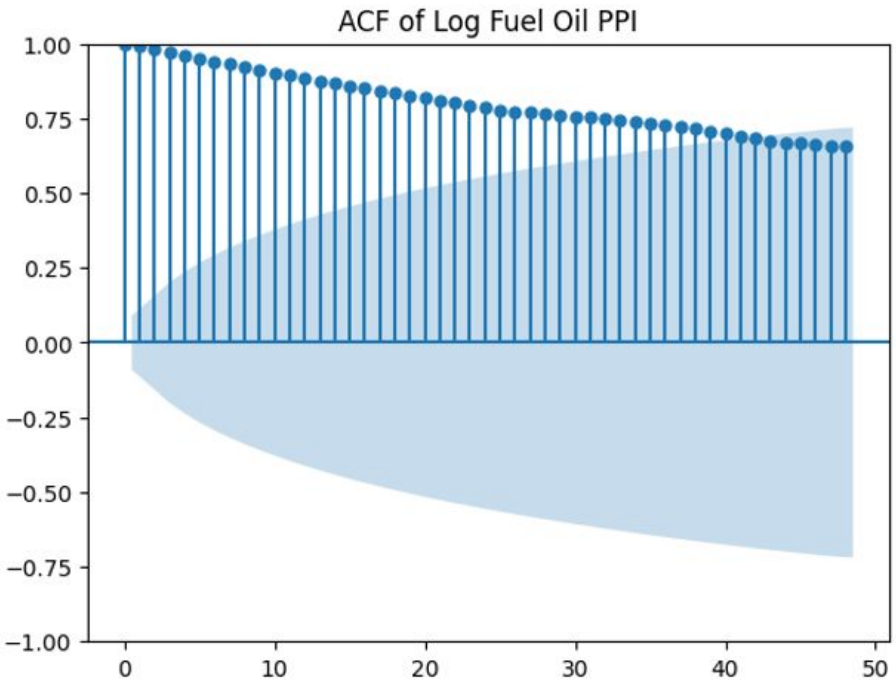
Correlation with PPI:

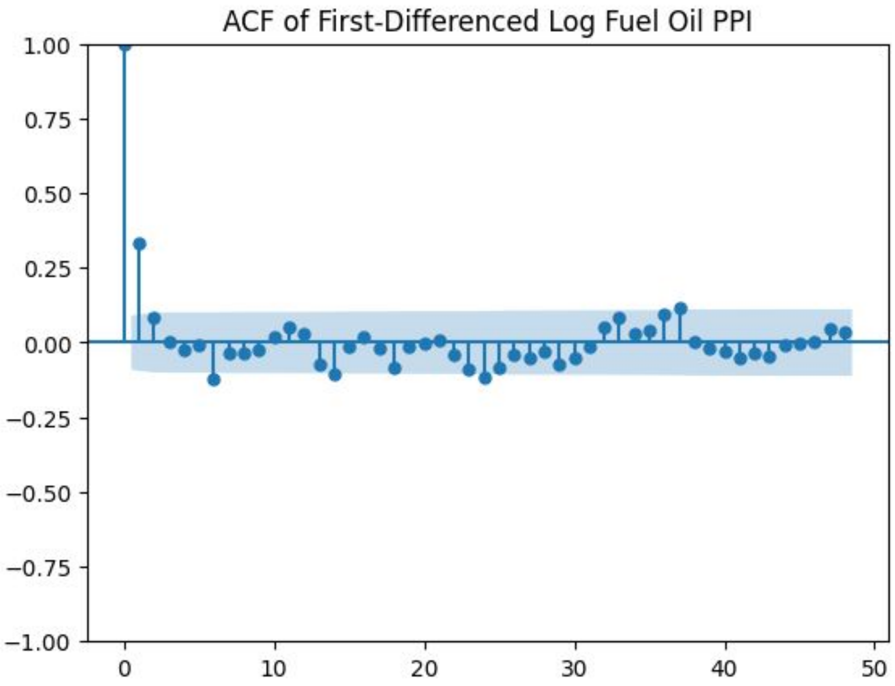
unemployment	0.003260
CPI	0.885557
Crude Oil (WTI) Price	0.938668
Industrial Production Index	0.826716

Final selected exogenous features: CPI, Crude Oil (WTI) Price, Industrial Production Index


Time Series Plot of Raw Fuel Oil PPI
fuel-oil
1
2
3
4
5
6
1990
1995
2000
2005
2010
2015
2020
2025


Time Series Plot of Log Fuel Oil PPI
log_oil
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
1990
1995
2000
2005
2010
2015
2020
2025





Original Series Stationarity Test:

ADF Statistic: -1.4299, p-value: 0.5679

Differenced Series Stationarity Tests:

ADF Statistic: -15.2333, p-value: 0.0000

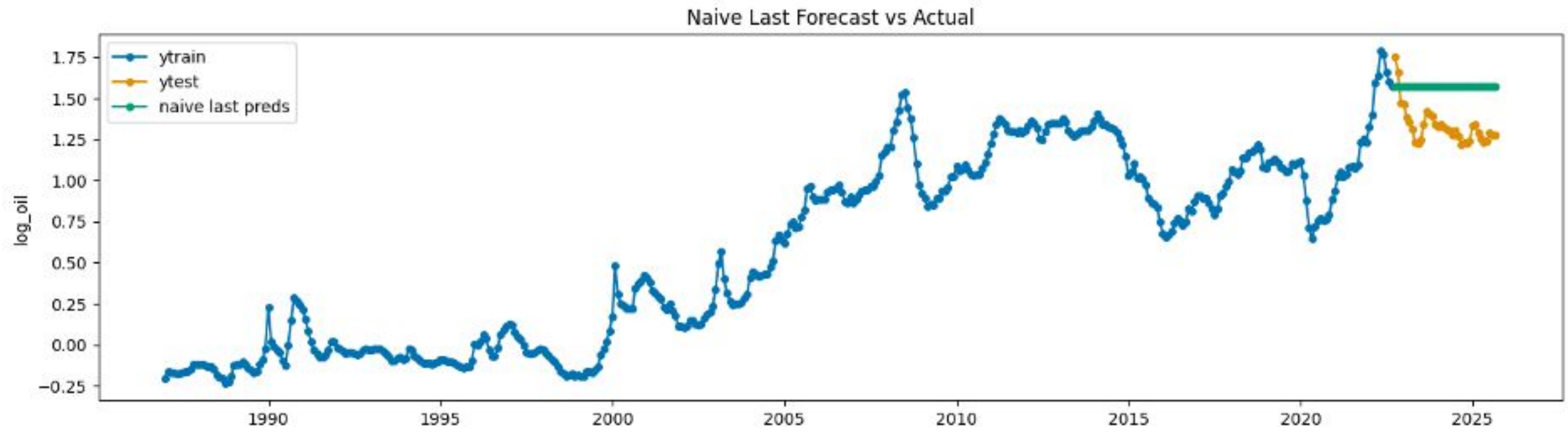
03

Models and Methods

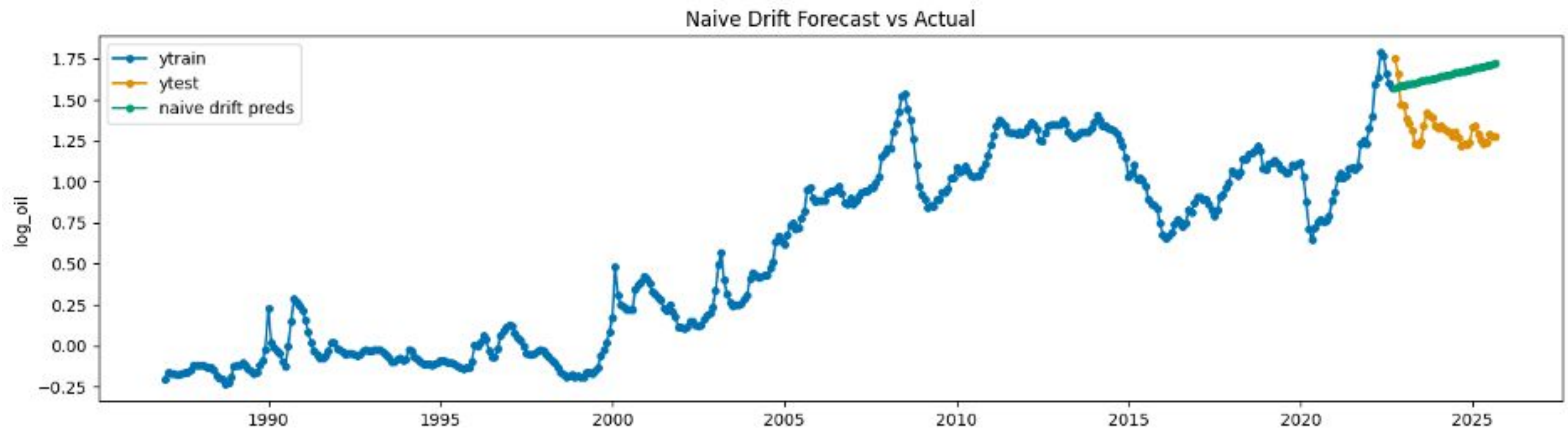
Benchmark Models

- Naive Last Value
- Naive Drift
- Naive Seasonal

Naive Last Value

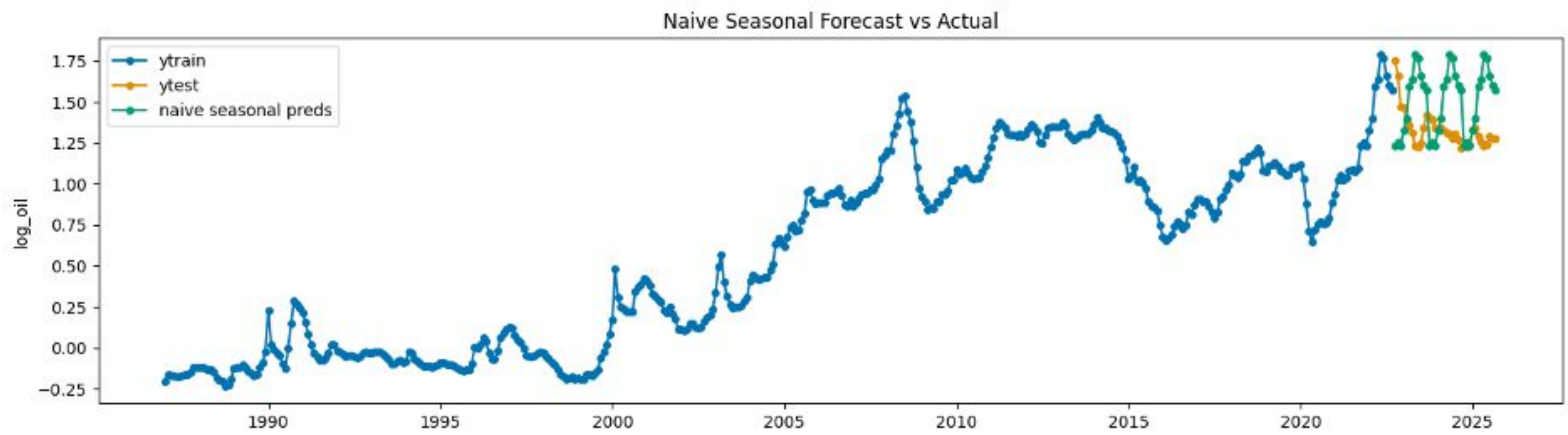


Naive Drift



Naive Drift Forecast vs Actual
ytrain
ytest
naive drift preds
log_oil
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
1990
1995
2000
2005
2010
2015
2020
2025

Naive Seasonal

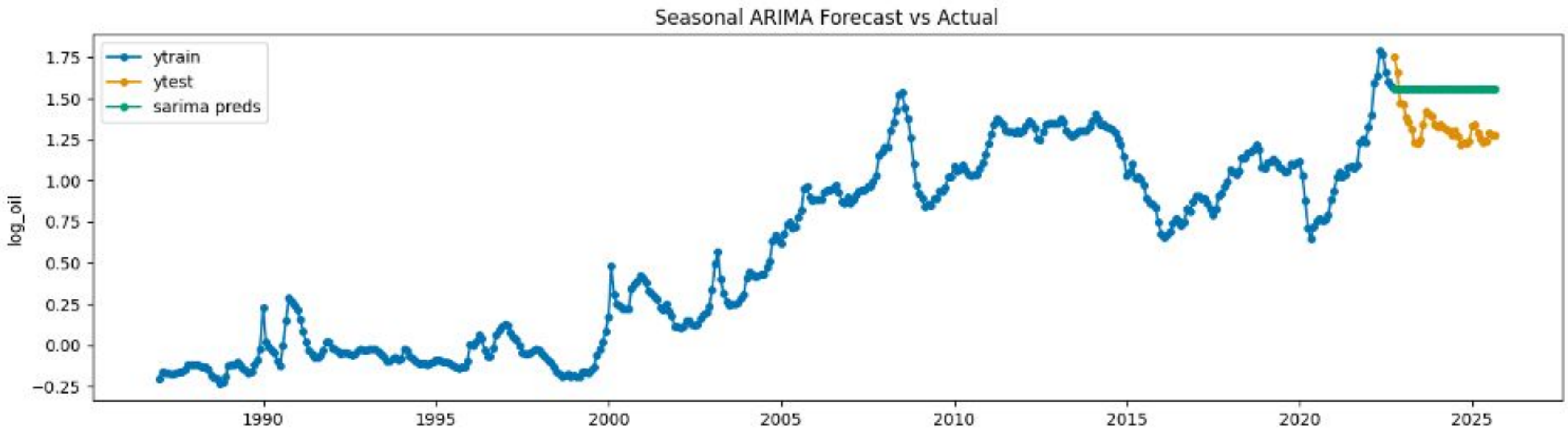


Classical Models

- **SARIMA**: Seasonal ARIMA models were fit using automated hyperparameter selection via AutoARIMA.

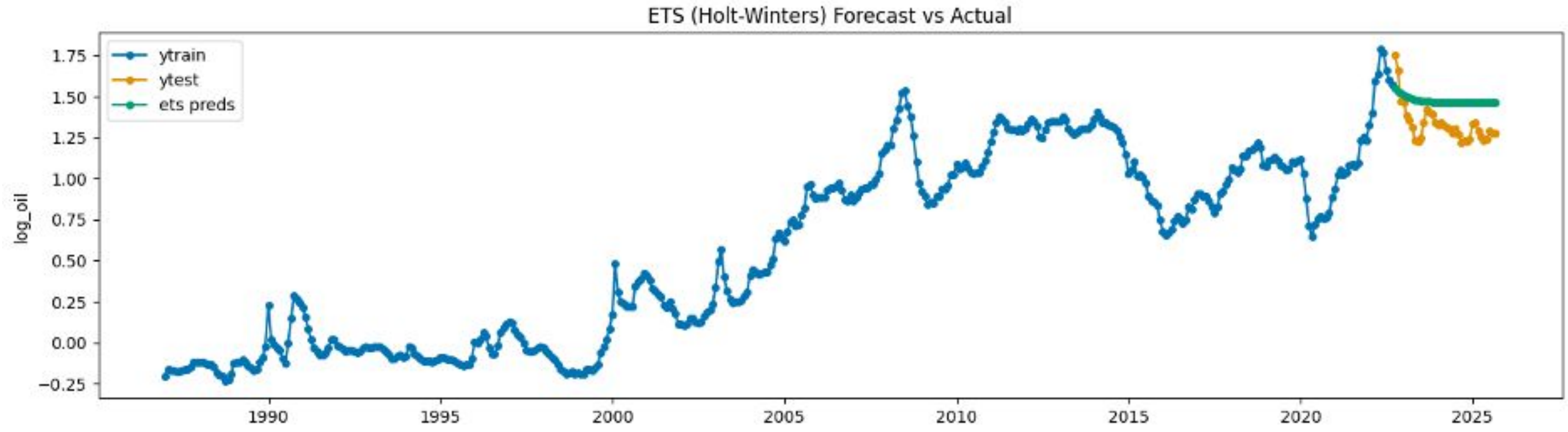
- **ETS**: Automatic Holt-Winters exponential smoothing using AutoETS

SARIMA

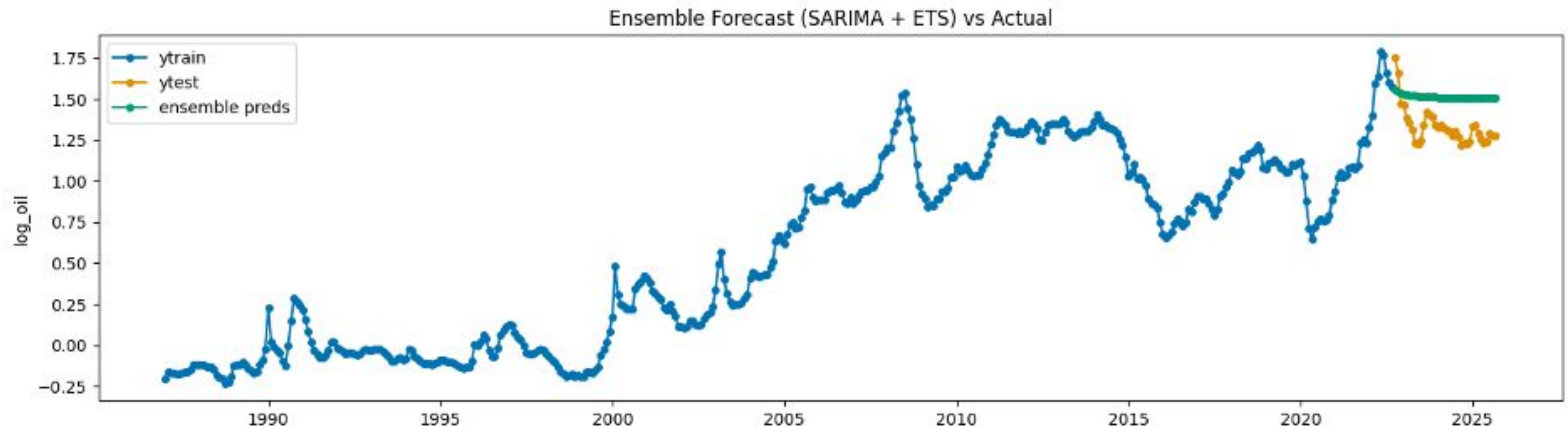


Seasonal ARIMA Forecast vs Actual
ytrain
ytest
sarima preds
log_oil
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
1990
1995
2000
2005
2010
2015
2020
2025

Exponential Smoothing (ETS)



Ensemble Model



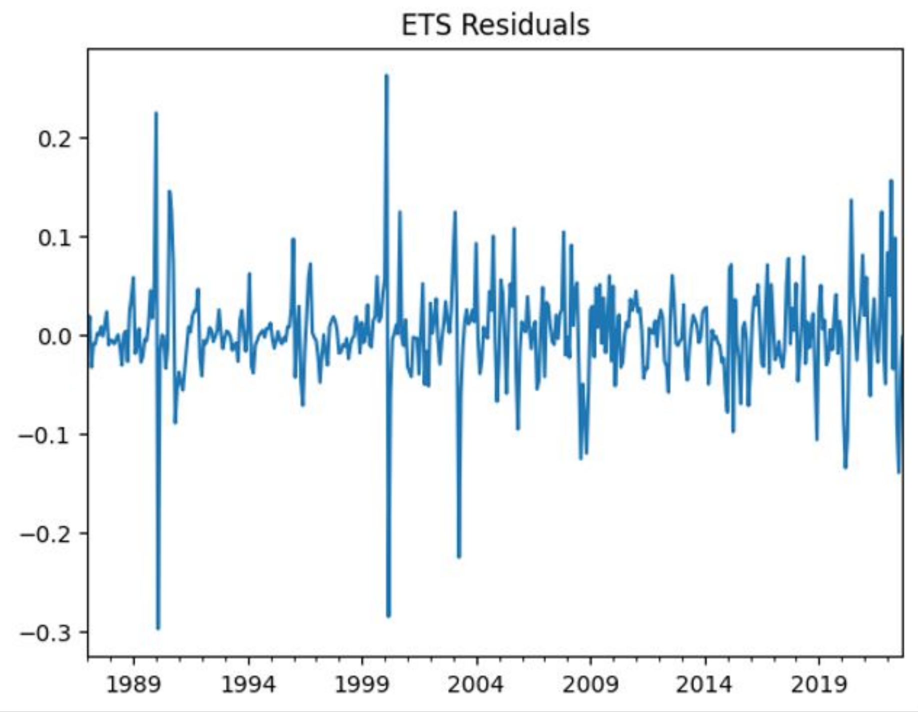
Hybrid ETS + XGBoost Model

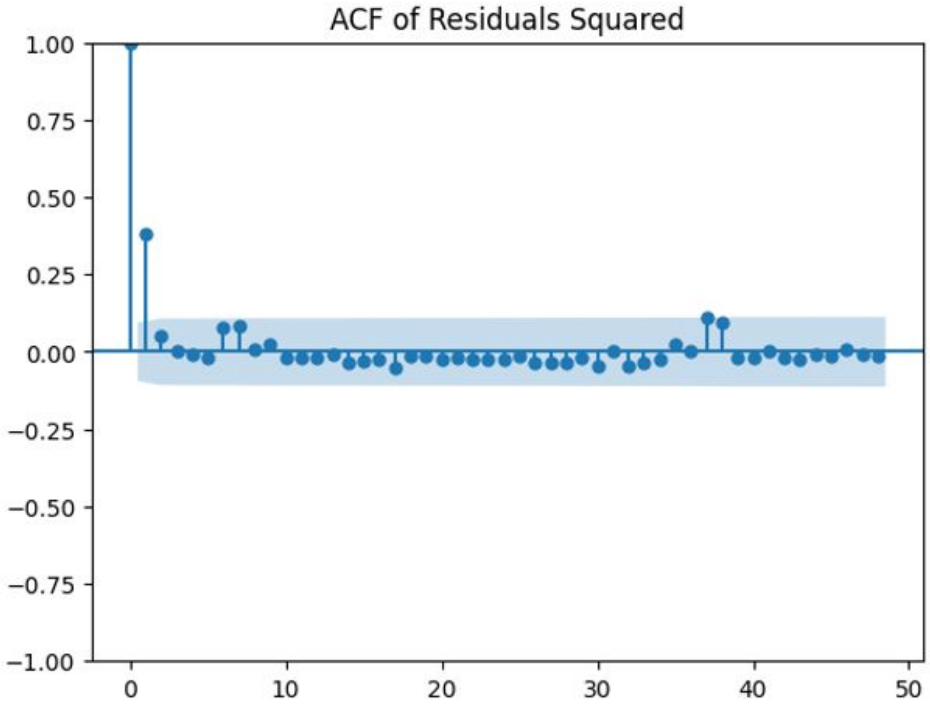
Step 1 – Baseline Forecast Using ETS Model

	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

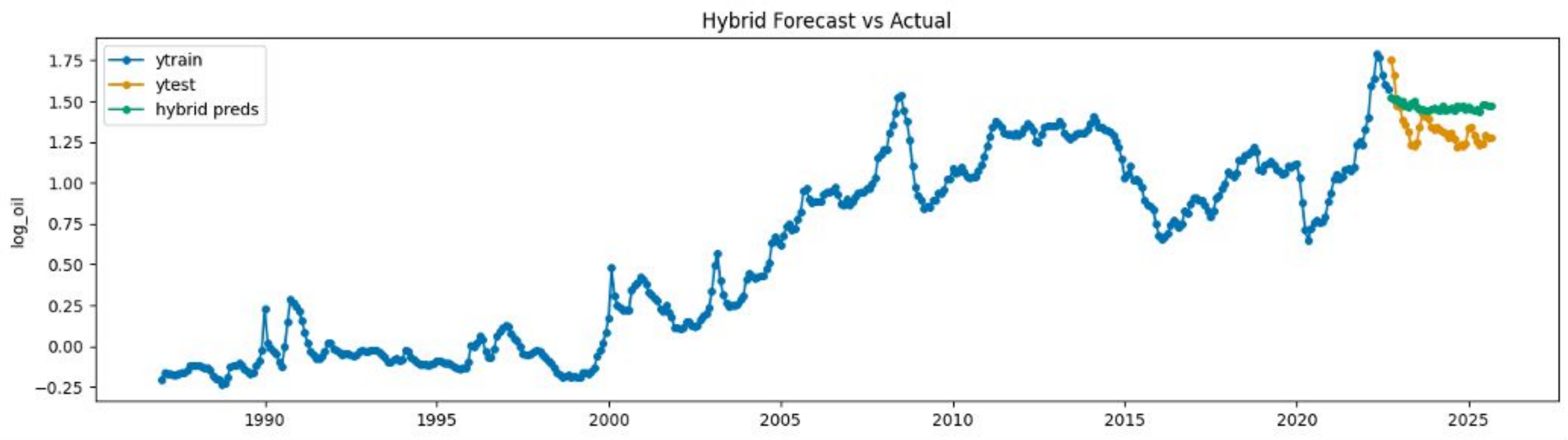
- Forecastability checking
- XGBoost Model


ETS Residuals
0.2
0.1
0.0
−0.1
−0.2
−0.3
1989
1994
1999
2004
2009
2014
2019


ACF of Residuals Squared
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50

Step 3 - Final Hybrid Forecast

ETS PPI forecast + XGBoost residual forecast



04

Results and Interpretation

Model Evaluation Method:
Mean Absolute Percentage Error (MAPE)

	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

Conclusion

- Classical time-series models such as ETS and ARIMA for seasonal economic forecasting are effective
- Incorporating macroeconomic drivers through machine learning provides improvements in predictive accuracy

Further Analysis

- Expand the set of macroeconomic features
- Explore alternative machine learning techniques for residual forecasting, such as neural networks
- Optimize ensemble strategies
- Incorporate probabilistic forecasts to quantify uncertainty

Thank you!