Cracking the Climate-Conscious Hard Commodities Code: Discovering Their True Value

Florence Chen Xu

Catalogue

Valuing Hard Commodities in Dynamic Climate

Compare and forecast hard commodities using multiple linear regression and machine learning models, incorporating insights from previous chapters.

Methodology

✣ **Data Selection**

- *Major Hard Commodities (Dependent Variables)*
  1. WTI Crude Oil
  2. Copper
  3. Henry Hub Natural Gas
- *Feature Selection (Independent Variables)*
  - Futures Close Price & Trading Volume: Copper, Aluminum, Gold, WTI Crude Oil, Henry Hub Natural Gas
  - Volatility Index (VIX)
  - USD/EUR Exchange Rate
  - Electricity Daily Generation Per Energy Source
  - Electricity Power Operations - Total Interchange
- *Timeframes:* May 2024 - November 2024

✣ **Data Preprocessing**

- Data was split with 30% reserved for testing, maintaining chronological order.
- MinMax scaling was applied consistently across all models.

✣ **Analysis Methods**

- Multiple Linear Regression
- Machine Learning Model
  - XGBoost
  - Long Short-Term Memory (LSTM)

✣ **Data Source**

- U.S. Energy Information Administration (EIA)
- Yahoo Finance
- Alpha Vantage

✣ **Limitation**

- Different types of models could limit the power of comparison.
- The choice of splitting and scaling methods can impact results.
- Variations in library versions (e.g., scikit-learn, XGBoost) can affect outcomes.
- All findings and recommendations are based on a selected period and features. The results may vary with different datasets, market conditions, or feature selections.

Key Findings

- Strong correlations among supply chain variables, such as weekly imports and refiner production, introduce *multicollinearity risks*. Diversifying feature selection enhances model robustness, but hidden multicollinearity — such as between supply and demand factors — must be carefully managed.
- Improper *scaling* can lead to data leakage, undermining predictive reliability. Maintaining proportionality between dependent and independent variables is crucial for preserving model performance
- Enabling *shuffling* during data splitting introduces volatility that benefits XGBoost. For short-term price forecasting, adjusting the sample size by incorporating seasonal patterns (e.g., based on the same quarter) can achieve comparable effect, improving trend recognition and ensuring balanced training and testing datasets.
- *Assumption violations* in multiple linear regression models affect absolute validity. However, analyzing these violations provides deeper insights into variable interactions, offering a more comprehensive financial perspective on market behavior.
- LSTM
  - While extensive *tuning* enhances statistical accuracy, it risks over-smoothing critical climate-related fluctuations. Increasing past data points to estimate current values makes LSTM less adaptive to short-term market shifts compared to multiple regression, but it remains valuable for capturing long-term trends.
  - Determining an appropriate number of past data points (*timesteps*) is key to recognizing seasonal price patterns in hard commodities. Additionally, effective *outlier* handling is vital, as removing them can disrupt sequential dependencies and skew predictions

Introduction

The concept of **equity** takes on different meanings across disciplines—whether in *law, finance,* or *climate change*—yet it consistently emphasizes the importance of balance. In *law,* equity ensures fairness by providing flexibility when strict legal rules fall short. In *finance,* it represents ownership and value. In *climate* discussions, equity focuses on the fair distribution of resources and responsibilities. Similarly, this analysis seeks to balance key factors affecting commodity prices, aiming to enhance forecasting reliability in an increasingly uncertain environment.

PART 1
Bridging Equity Valuation Components

Chapter I explored sustainable growth in climate change, emphasizing valuation factors like workforce distribution, geographic concentration, and demand patterns. While not previously discussed, **Total Factor Productivity (TFP)** is crucial for industries reliant on hard commodities, as renewable energy adoption impacts mineral production and price fluctuations.

Since hard commodity prices are a primary driver of revenue for energy and mining companies, the analysis shifts from evaluating growth sustainability in the denominator to assessing revenue determinants in the numerator. **Chapter II** examined the interconnectedness of hard commodity prices with market-relevant factors, such as the Consumer Price Index (CPI), to better understand their relationship with profitability. Building on this foundation, **Chapter III** takes a more direct approach to revenue forecasting, employing multiple regression and machine learning models to refine commodity price predictions.

Chapter II
The Interconnectedness

Chapter III
Valuing Hard Commodities in Dynamic Climate

$$\frac{FCFF}{WACC - g}$$

Chapter I
The Sustainable Growth Rate

PART 2
Reviewing Forecasting Approaches

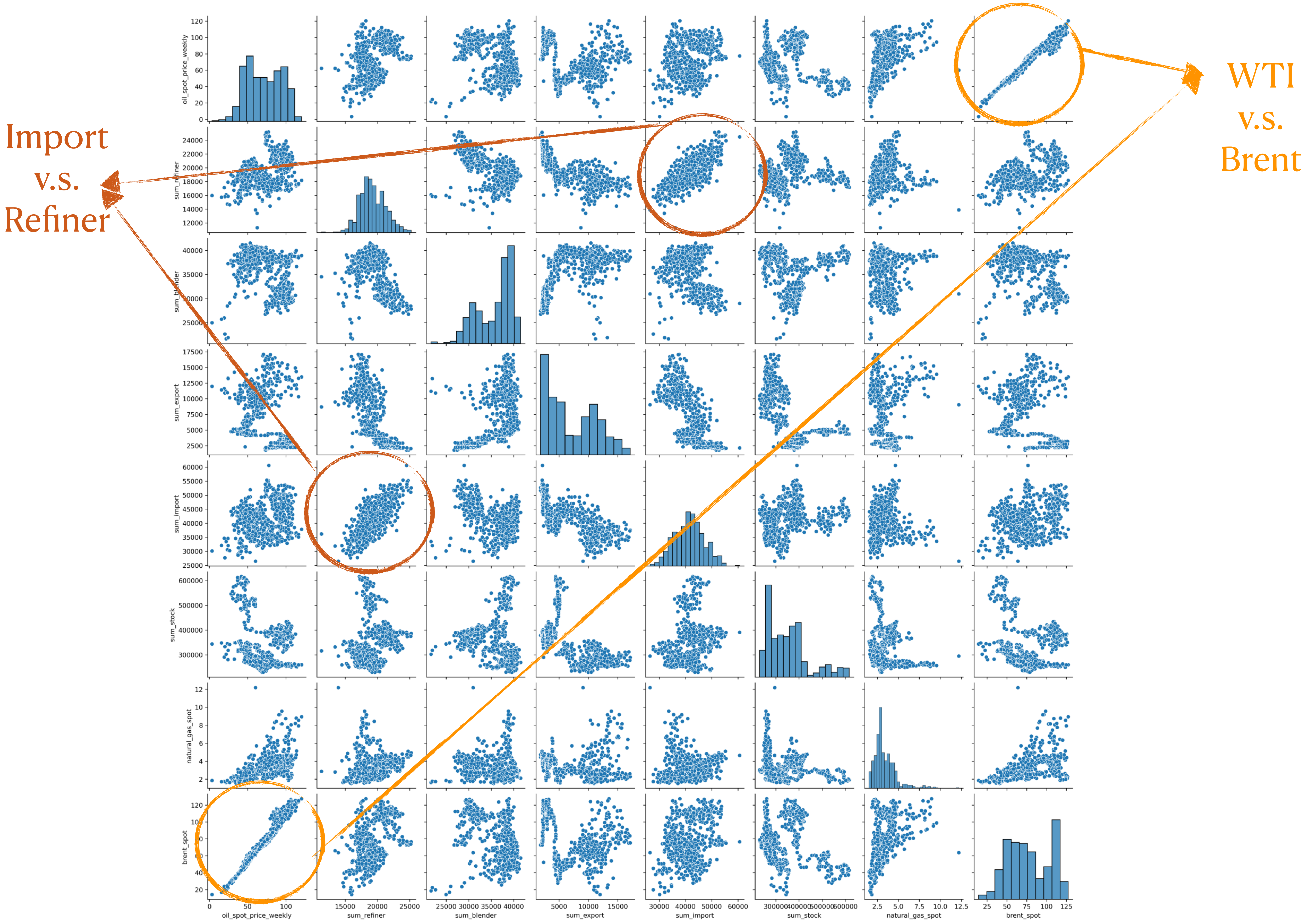
In my previous project, ***Crude Oil Within the Context of an Oligopoly Market***, XGBoost was utilized for crude oil forecasting, but it lacked detailed statistical analysis. To address this, a multiple linear regression model will be applied to better understand the relationships between key factors and their coefficients. While crude oil prices may not exhibit a strictly linear relationship with these factors, this approach will provide a foundational understanding before advancing to more complex models like XGBoost and Long Short-Term Memory (LSTM).

Multicollinearity

Strong correlations among supply chain features—e.g., a high positive correlation between weekly imports and refiner production—necessitates more diversified feature selection for robust modeling

Variance Inflation Factor

	Variable	VIF
0	Refinery Output	187.140106
1	Blended Fuel Output	111.672283
2	Total Exports	12.607574
3	Total Imports	143.640652
4	Inventory Levels	27.035439
5	Natural Gas Price	10.964020
6	Brent Crude Price	20.129991



XGBoost vs. Multiple Regression

XGBoost, which is free from strict assumptions, handles extreme values better, making it suitable for hard commodities amid climate uncertainties. However, its reliance on preprocessing and data selection introduces risks, as dataset scaling before splitting and the use of shuffling led to potential data leakage, compromising predictive reliability.

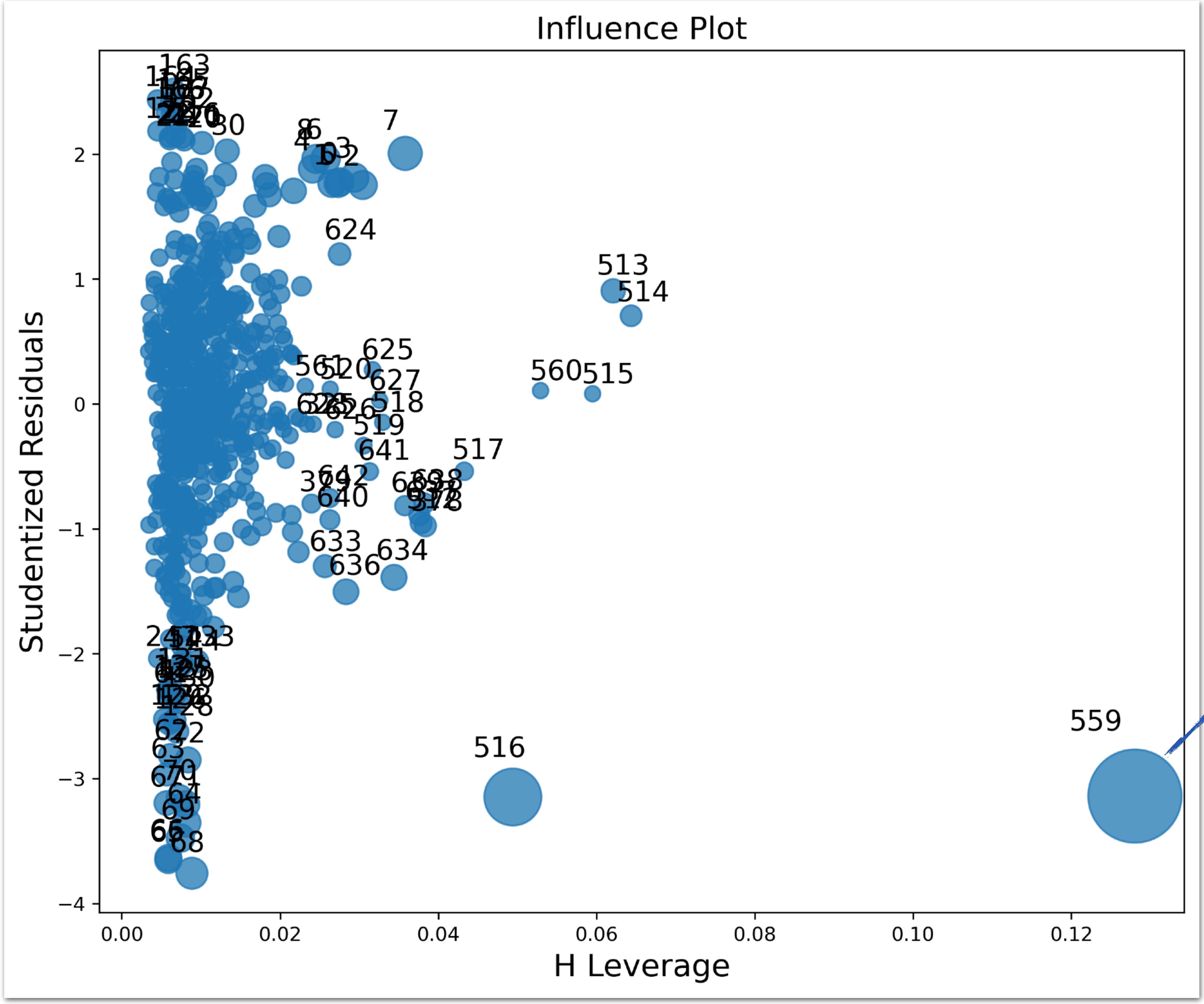
Model/ Metrics	XGBoost		Multiple Linear Regression
	Scenario A (All features included)	*Scenario B (Excluding highly correlated feature – Crude Oil Brent*	
RMSE	2.974	5.063	3.954
R-squared	0.983	0.952	0.969

Source: XGBoost results from "Crude Oil Within the Context of an Oligopoly Market". Available at: https://github.com/florencex5/Crude_Oil_Finance_Project.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     oil_spot_price_weekly   R-squared:                       0.969
Model:                               OLS   Adj. R-squared:                  0.969
Method:                    Least Squares   F-statistic:                     3087.
Date:                   Tue, 11 Feb 2025   Prob (F-statistic):               0.00
Time:                           14:56:20   Log-Likelihood:                -1941.6
No. Observations:                    695   AIC:                             3899.
Df Residuals:                        687   BIC:                             3935.
Df Model:                              7
Covariance Type:               nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept            -0.2757      3.485     -0.079      0.937      -7.118       6.567
sum_refiner          -0.0002      0.000     -1.385      0.166      -0.000    7.42e-05
sum_blender           0.0002   5.01e-05      3.265      0.001    6.52e-05       0.000
sum_export        -8.488e-06    7.7e-05     -0.110      0.912      -0.000       0.000
sum_import        -2.057e-05   4.45e-05     -0.462      0.644      -0.000    6.68e-05
sum_stock          2.914e-06   2.47e-06      1.179      0.239   -1.94e-06    7.77e-06
natural_gas_spot      1.3390      0.137      9.805      0.000       1.071       1.607
brent_spot            0.8290      0.008     99.054      0.000       0.813       0.845
==============================================================================
Omnibus:                       52.963   Durbin-Watson:                   0.160
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               88.206
Skew:                          -0.535   Prob(JB):                     7.02e-20
Kurtosis:                       4.379   Cond. No.                     8.47e+06
==============================================================================
```

Outlier Impact

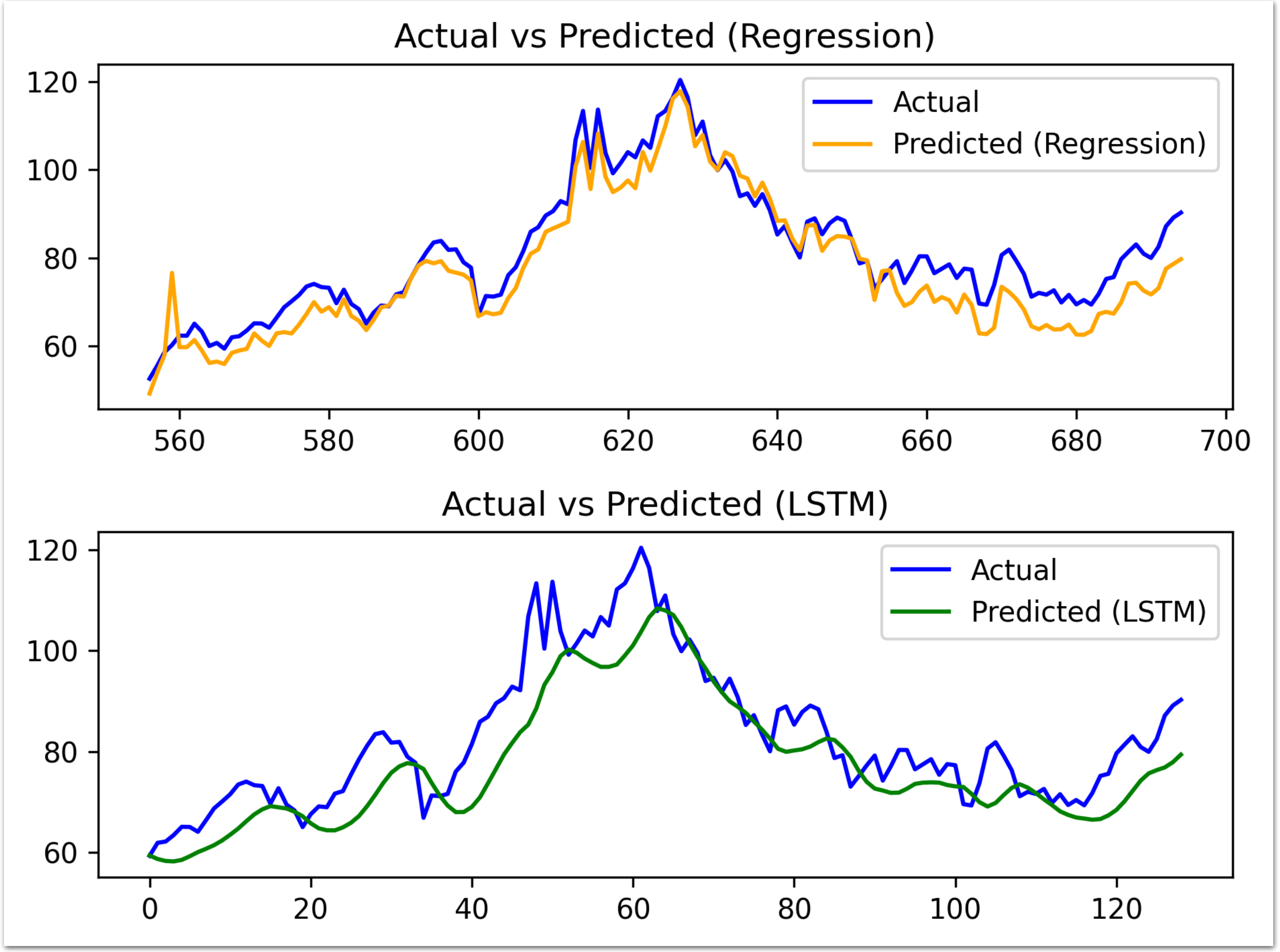
In hard commodity forecasting, outliers can have lasting effects. In LSTM models, removing outliers disrupts sequential patterns, potentially leading to inaccurate forecasts. Proper handling of outliers—whether through removal or winsorization—requires careful consideration of industry-specific factors.



Observation #559 exceeds the Cook's D threshold, suggesting high influence

LSTM Limitations for Hard Commodities

Extensive tuning improves accuracy but may over-smooth critical climate fluctuations. Multiple regression, depending on features and timeframe, offers more responsive predictions, while LSTM excels in capturing long-term trends.



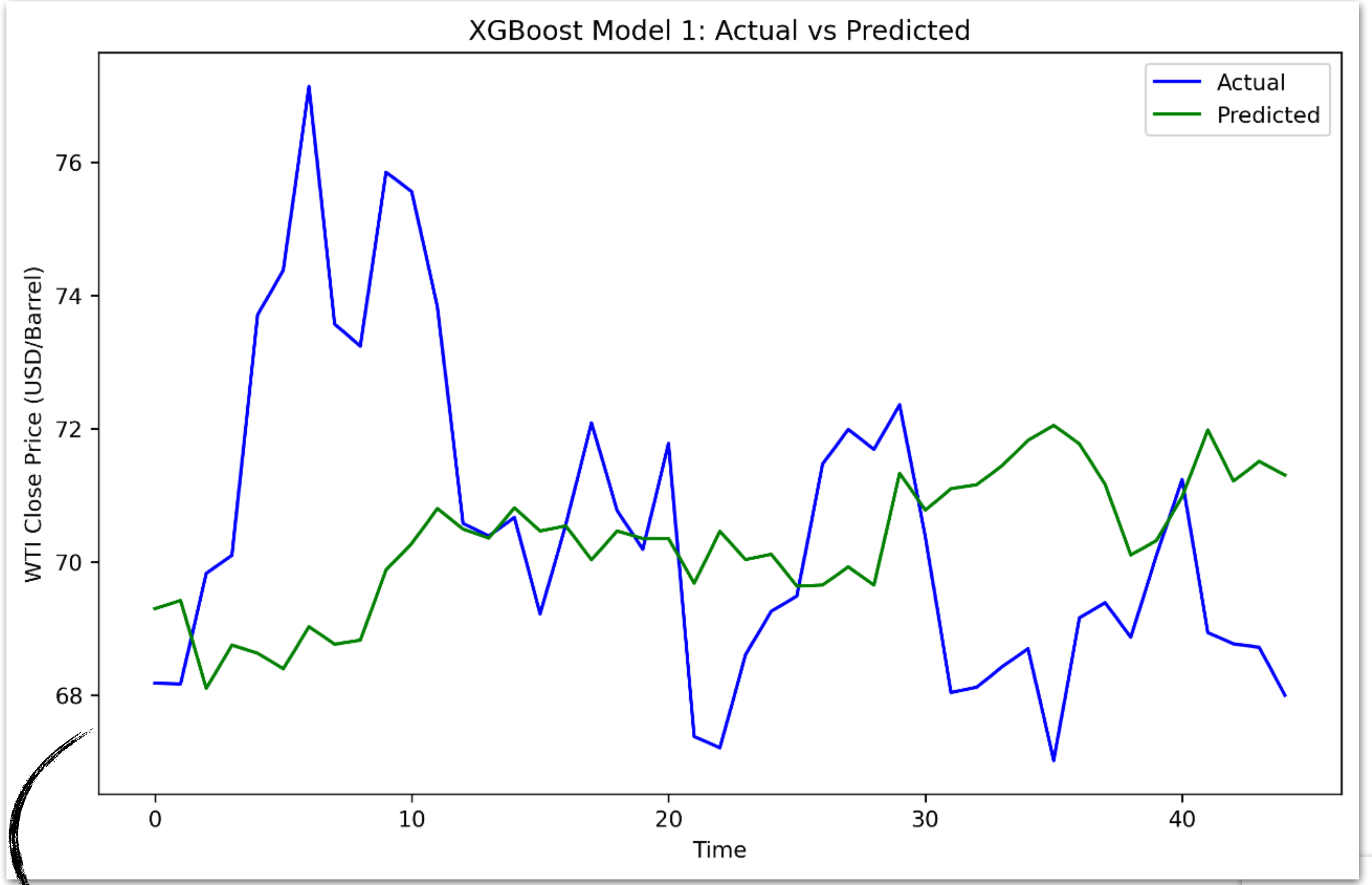
PART 3
Commodities Valuation Model

For more details on the analysis and model implementation, please refer to the corresponding Jupyter notebook.

- WTI Crude Oil
- Copper
- Henry Hub Natural Gas

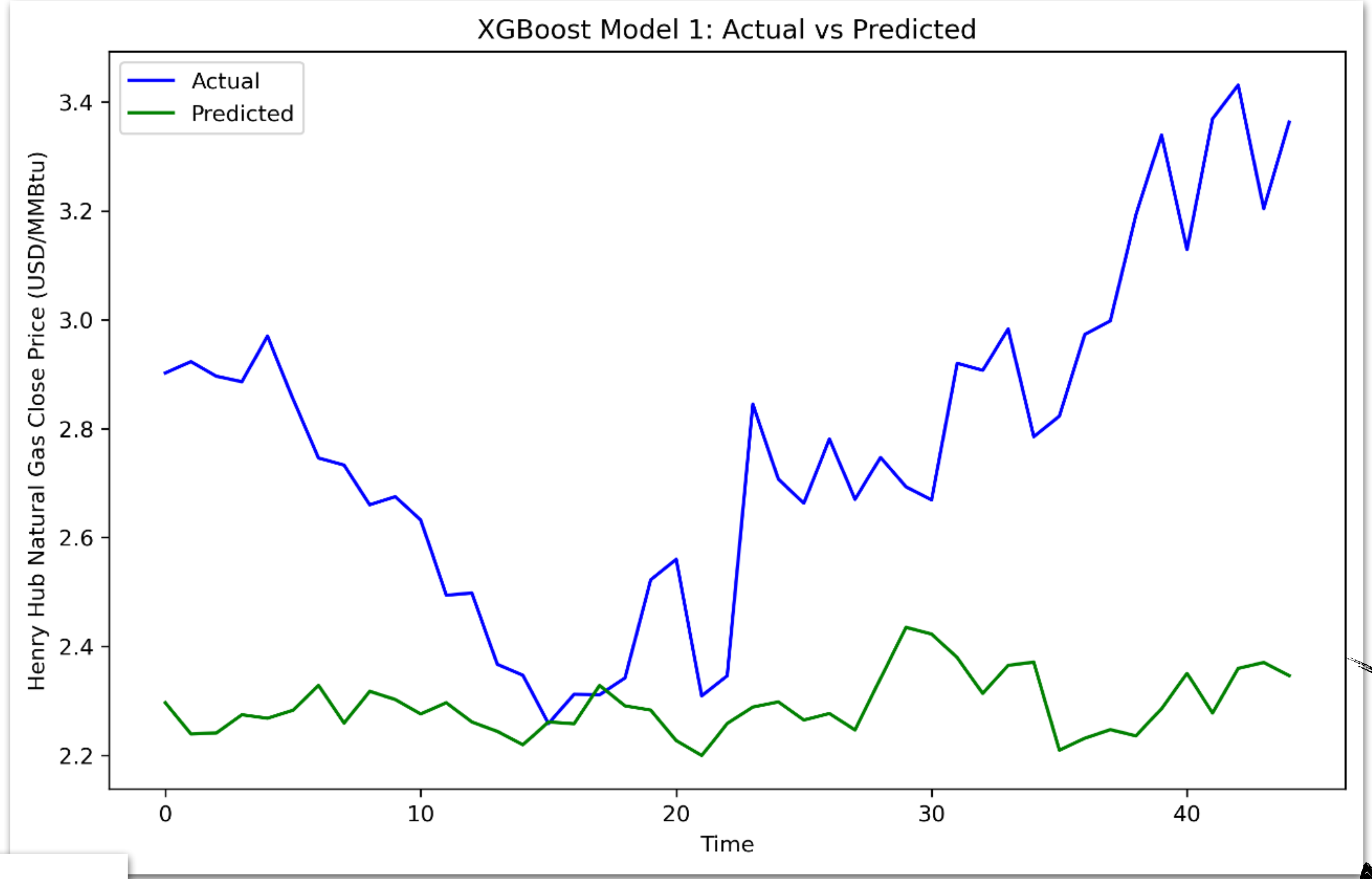
Overfitting and Volatility Sensitivity

XGBoost performed well in training but showed significant test data decline, indicating overfitting. Its tendency to smooth trends rather than capture extreme values limits its reliability for forecasting hard commodities. Prediction inconsistencies also arose from scale disparities between dependent and independent variables, such as futures prices versus trading volume.

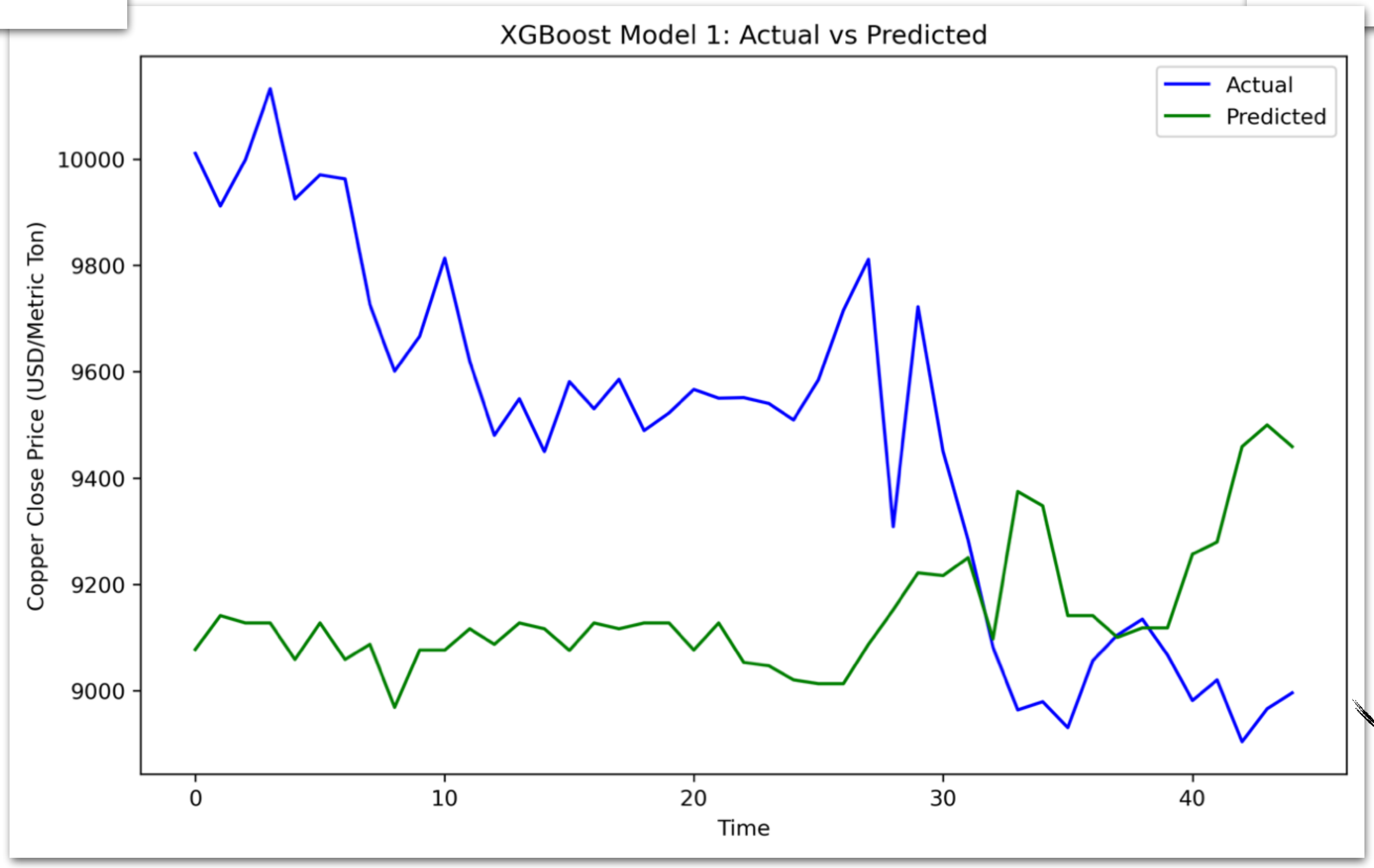


WTI Crude Oil

- *Model trained with tuned hyperparameters*
- *Data split without shuffling*



Henry Hub Natural Gas



Copper

Multicollinearity in Forecasting

Despite high VIF values indicating multicollinearity (e.g., USD/EUR at 1873 for natural gas), these variables consistently ranked among the top five in XGBoost feature importance across all selected hard commodities, including WTI and copper, highlighting their significance in forecasting.

Variance Inflation Factor (WTI)

	Variable	VIF
0	Gold Closing Price	1354.720371
1	Gold Trading Volume	1.248844
2	WTI Trading Volume	13.194334
3	Natural Gas Closing Price	94.289186
4	Natural Gas Trading Volume	19.214697
5	Volatility Index (VIX)	38.550132
6	USD to EUR Exchange Rate	927.251193
7	Electricity Generated from Oil	3.979410
8	Electricity - Total Interchange	1.757855
9	Month	91.835797
10	Day	7.163156

Variance Inflation Factor (Natural Gas)

	Variable	VIF
0	Gold Closing Price	1432.401157
1	Gold Trading Volume	1.248177
2	WTI Closing Price	1044.486573
3	WTI Trading Volume	14.240641
4	Natural Gas Trading Volume	18.795562
5	Volatility Index (VIX)	33.693159
6	USD to EUR Exchange Rate	1872.985248
7	Electricity Generated from Natural Gas	98.099030
8	Electricity - Total Interchange	2.600768
9	Month	124.948054
10	Day	7.918195

Variance Inflation Factor (Copper)

	Variable	VIF
0	Copper Trading Volume	1.835063
1	Aluminum Closing Price	719.867843
2	Aluminum Trading Volume	1.427376
3	Gold Closing Price	2143.108260
4	Gold Trading Volume	1.582976
5	Volatility Index (VIX)	38.308404
6	USD to EUR Exchange Rate	831.260190
7	Electricity - Total Interchange	1.626976
8	Month	96.377837
9	Day	6.191957

Note: The comparison, based on VIF metrics, is limited by differing feature selections for each commodity, as their unique correlation patterns may affect the generalizability of the results.

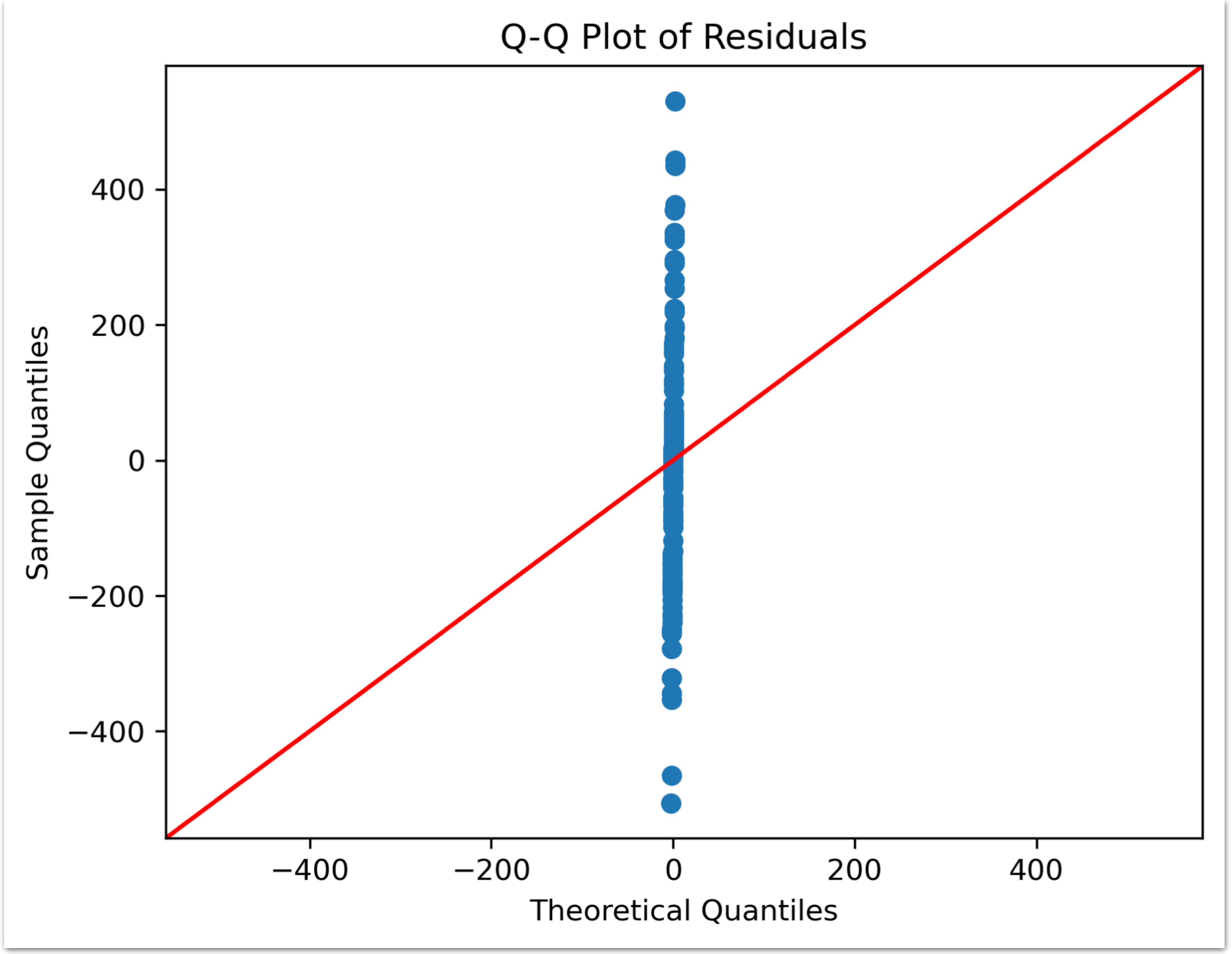
RMSE Comparison

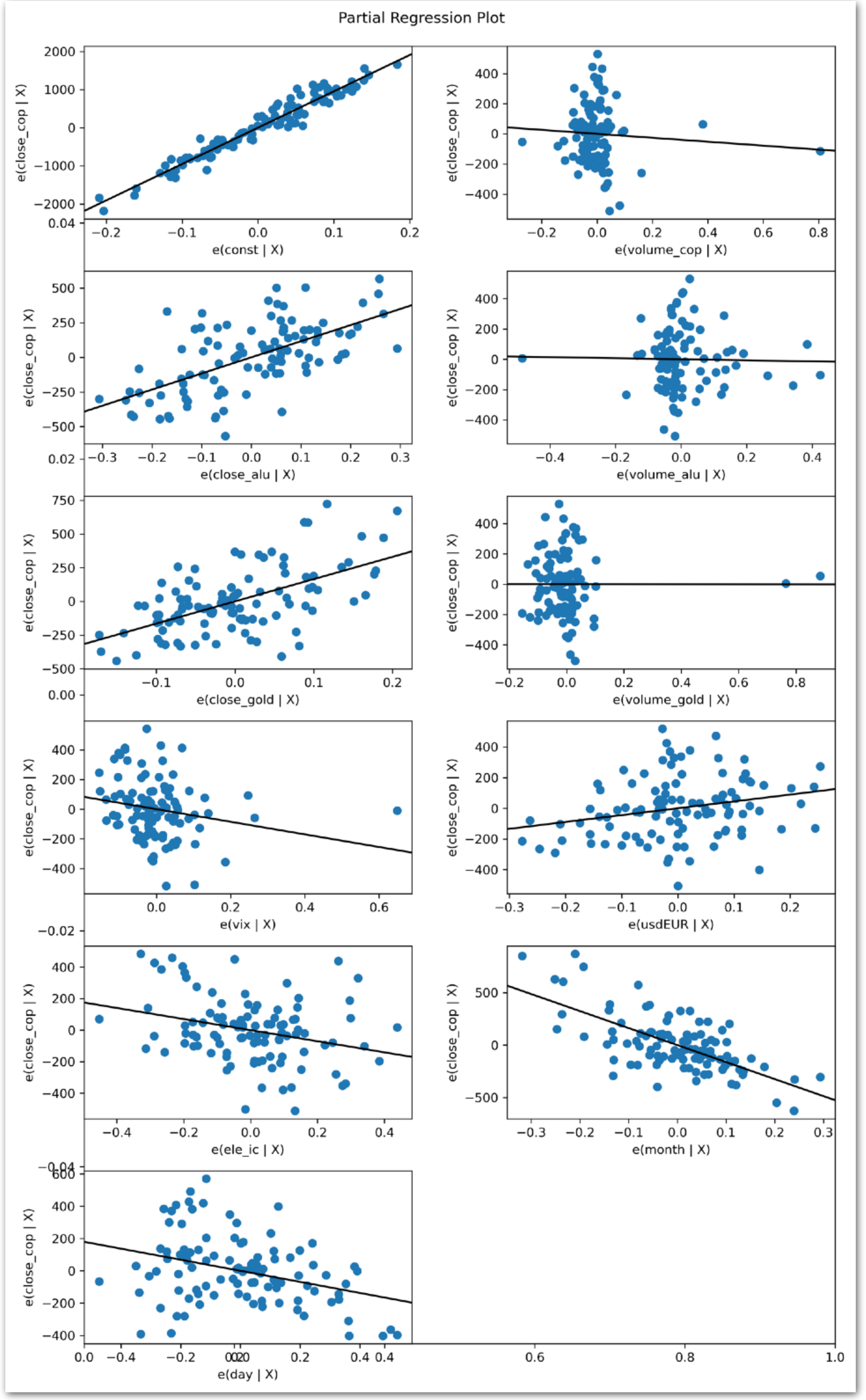
XGBoost performed well in training but showed significant test data decline, indicating overfitting. Its tendency to smooth trends rather than capture extreme values limits its reliability for forecasting hard commodities. Prediction inconsistencies also arose from scale disparities between dependent and independent variables, such as futures prices versus trading volume.

Model/ Hard Commodities	WTI Crude Oil	Henry Hub Natural Gas	Copper
Multiple Linear Regression	5.717	0.382	408.873
Multiple Linear Regression (Without Influential Points)	5.130	0.439	444.875
XGBoost - Optimal Model (Training Data)	1.560	0.075	516.998
XGBoost - Optimal Model (Test Data)	2.975	0.564	557.418
LSTM	7.826	0.405	560.901

Assumption and Prediction Quality

In climate-driven markets, assumption violations in multiple regression signal statistical issues but offer financial insights into variable relationships and market dynamics.



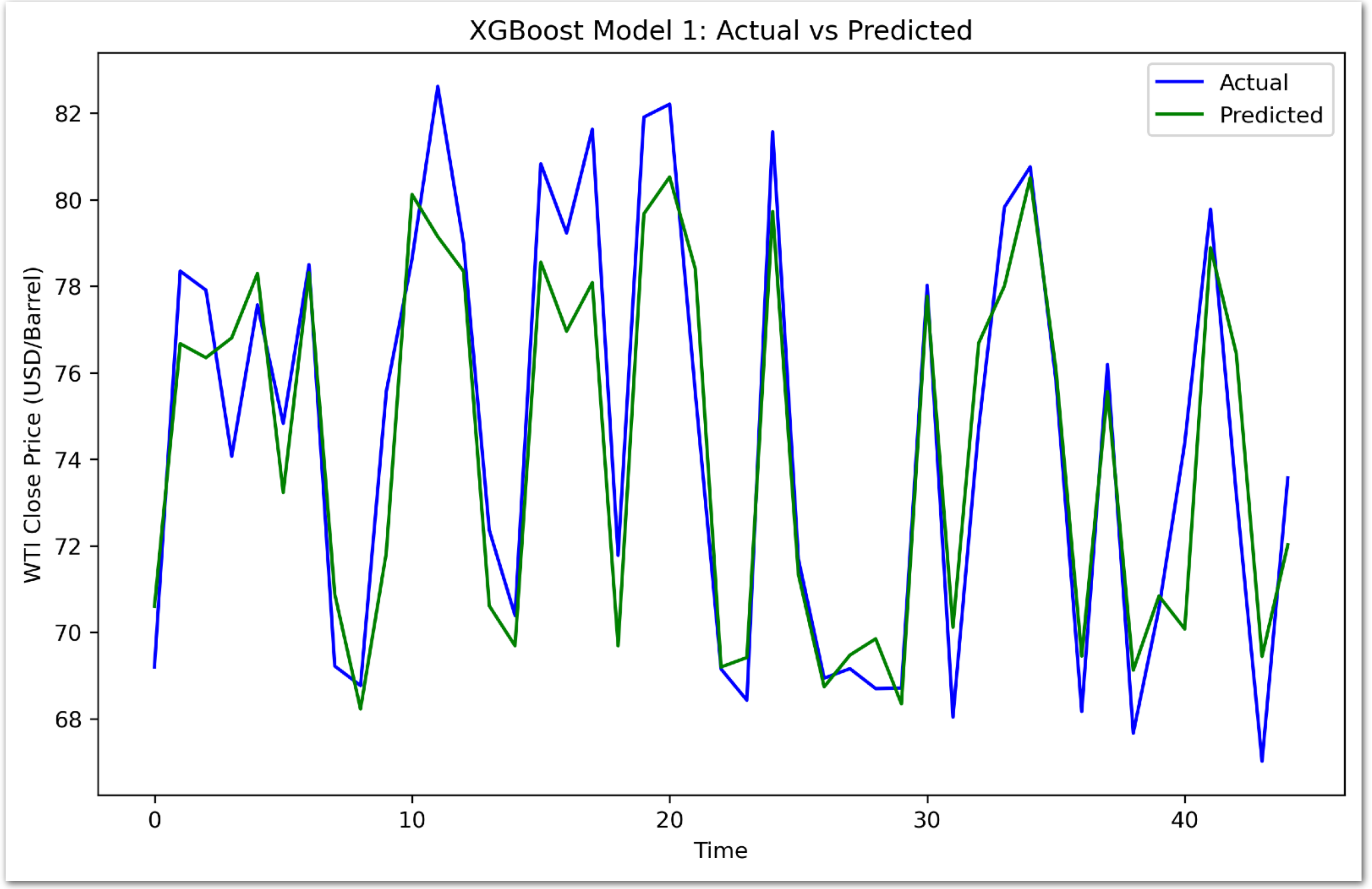


Source: Residual Analysis for Copper. Available at: https://github.com/florencex5/Hard_Commodities/blob/main/valuation_model_cop.ipynb

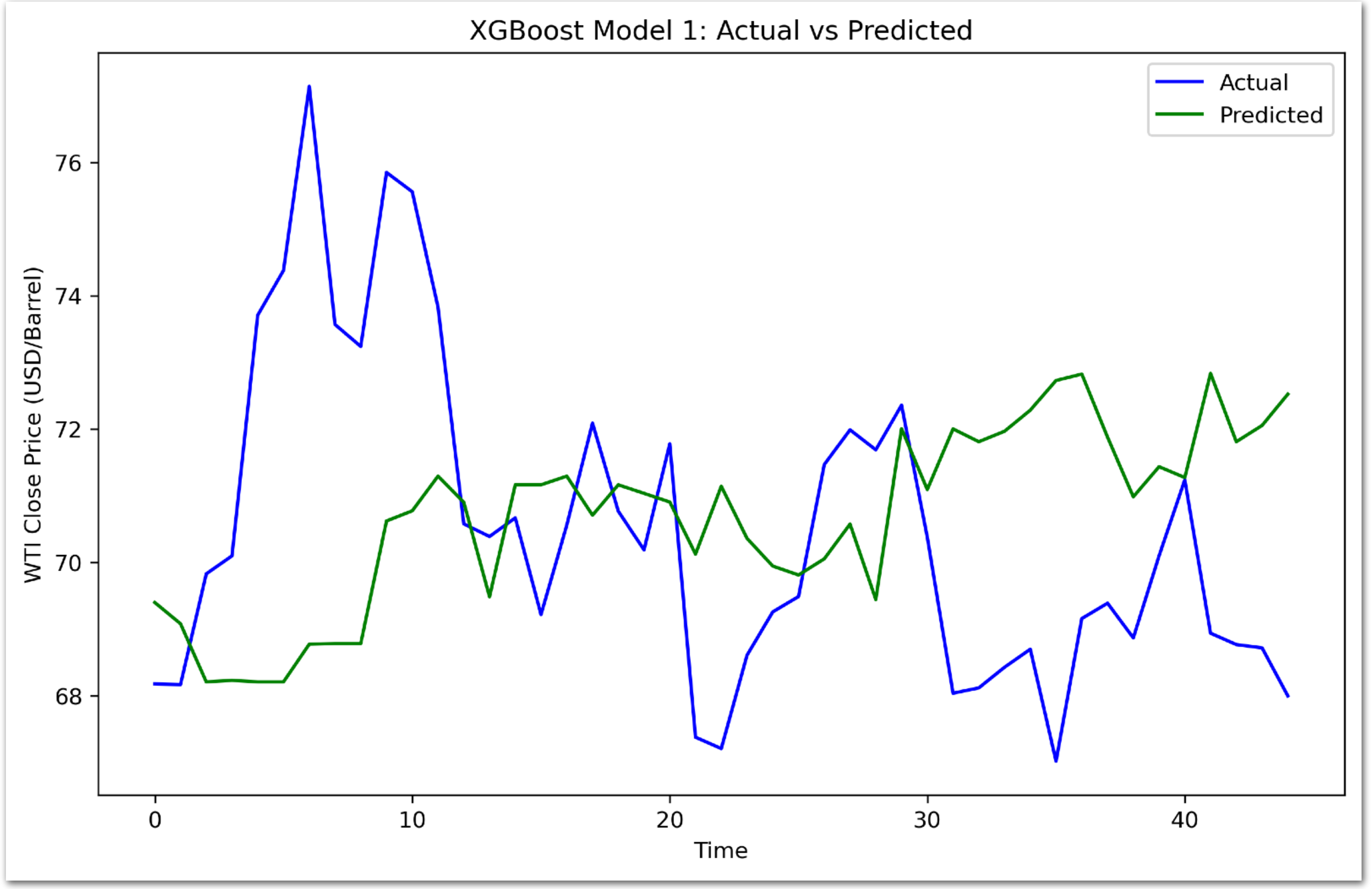
Recommendations

"Shuffling" vs. Seasonal Data Incorporation

- When splitting data for training and testing, shuffling introduces volatility and complexity, which can enhance XGBoost's performance. Alternatively, manually incorporating seasonal data can replicate this effect, restoring necessary fluctuations for better trend recognition. This approach not only improves prediction accuracy by creating meaningful "waves" in the data but also ensures a balanced expansion of the training and testing datasets.



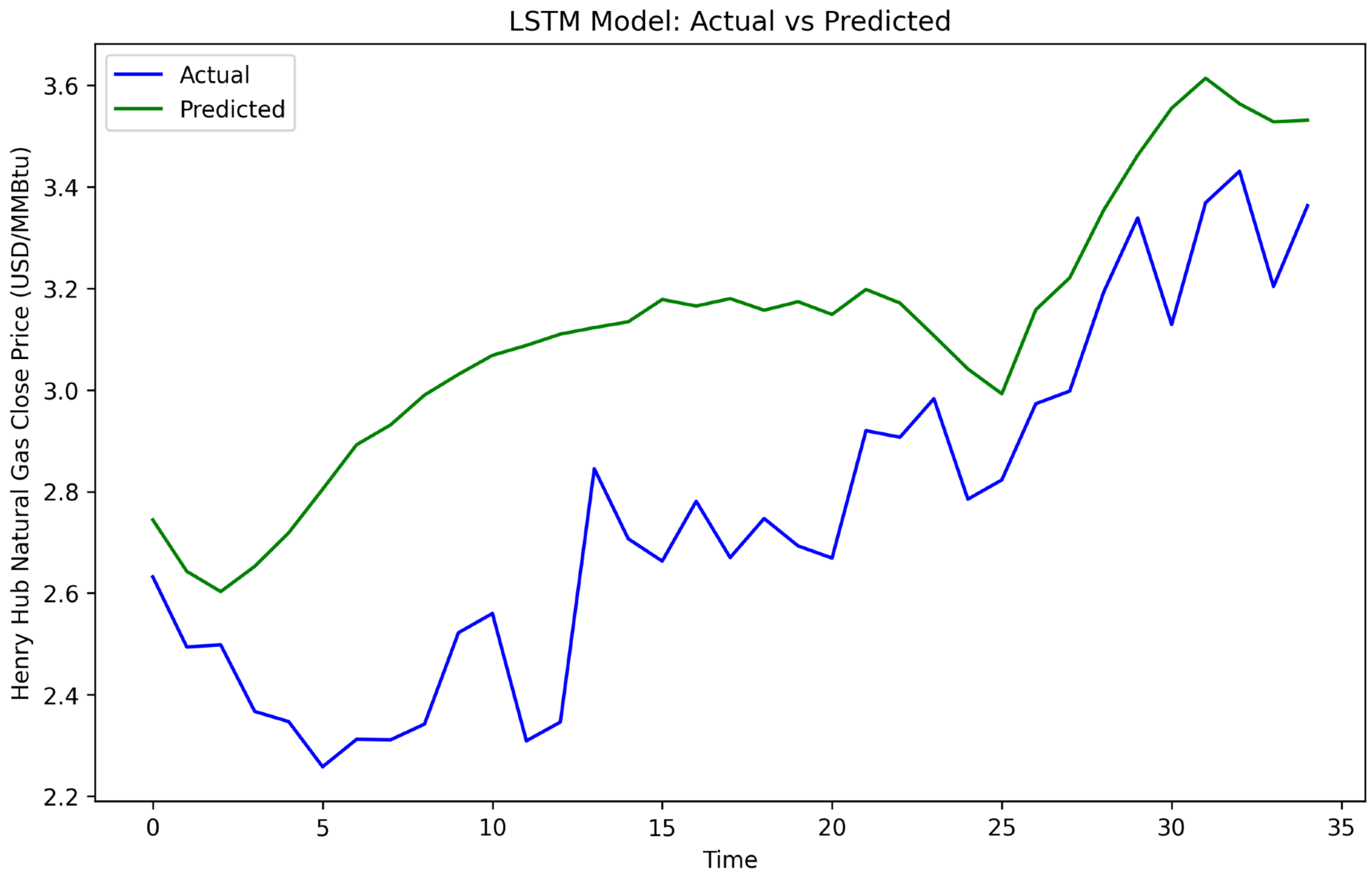
Shuffle = True



Shuffle = False

Recommendations

- **Addressing Data Imbalance:** Rather than standard scaling, proportionally adjusting dependent variables (e.g., multiplying by 100) can better align their magnitude with independent variables, minimizing discrepancies in scale and unit size.
- **Timesteps in Forecasting:** selecting the appropriate number of past data points—referred to as "timesteps"—is crucial for effective forecasting. This determines how much historical data the model considers when forecasting future values. For hard commodities with strong seasonal patterns—such as natural gas prices or gold trading volumes—adjusting timesteps helps capture these variations effectively, enhancing the model's predictive accuracy.



Reference

- PwC(2024 Apr). Climate risks to nine key commodities . Available at:https://www.pwc.com/gx/en/issues/esg/people-and-prosperity-at-risk.pdf
- Hiller J. (2024 Apr). How Big Data Centers Are Slowing the Shift to Clean Energy. WSJ. Available at: https://www.wsj.com/business/energy-oil/how-big-data-centers-are-slowing-the-shift-to-clean-energy-44ef4145?st=d697Wm
- Crum A. (2024 May). Exxon Mobil Takes On Climate Extremists. WSJ. Available at: https://www.wsj.com/articles/exxon-mobil-takes-on-climate-extremists-6948185e?mod=Searchresults_pos2&page=1
- December R. (2024 Jun). It's Going to Cost More to Stay Cool This Summer. WSJ. Available at: https://www.wsj.com/finance/commodities-futures/ac-electricity-bill-higher-summer-4b9a8edb?mod=saved_content
- Laffer A. & Moore S. (2024 Jun). Bidenomics, Also Known as MMT. WSJ. Available at: https://www.wsj.com/articles/bidenomics-also-known-as-mmt-modern-monetary-theory-spending-ccadee9c?mod=commentary_article_pos2
- Young L. (2024 Jun).Solar Energy Faces Cloudy Prospects on Warehouse Rooftops. WSJ. Available at: https://www.wsj.com/articles/solar-energy-faces-cloudy-prospects-on-warehouse-rooftops-abea837e?mod=saved_content
- Otis G. (2024 Aug). Hawaiian Electric to Settle Maui Wildfire Claims as Part of $4 Billion Deal. WSJ. Available at: https://www.wsj.com/business/hawaiian-electric-to-settle-maui-wildfire-claims-as-part-of-4b-deal-5c7d06f5?mod=saved_content
- Salama V. (2024 Nov). What Trump Can—and Can't—Do on Day One. WSJ. Available at: https://www.wsj.com/politics/elections/what-trump-canand-cantdo-on-day-one-a90a8799?mod=saved_content
- Morenne B. & Eaton C. (2024 Nov). Trump's Oil and Gas Donors Don't Really Want to 'Drill, Baby, Drill'. WSJ. Available at: https://www.wsj.com/business/energy-oil/trump-oil-gas-policy-drilling-donors-3438e99e?mod=saved_content
- Khan Y. (2024 Nov). Disclosure Isn't Just About Saving the Planet, It's a Business Necessity Now, Says CDP Chief. WSJ. Available at: https://www.wsj.com/articles/disclosure-isnt-just-about-saving-the-planet-its-a-business-necessity-now-says-cdp-chief-5577aaa9?mod=saved_content

THANK YOU