Crude Oil Gasoline Dynamics: Cointegration Forecasting for Energy Arbitrage

Mahima Masetty Nidhi Pareddy Suyog Mahale Zeel Patel

2025-03-12

Table of Contents

1. Introduction

The relationship between gasoline and crude oil prices has long been recognized as a crucial indicator in energy markets, primarily due to their inherent cointegration. This close relationship exists because Gasoline is a refined derivative of Crude Oil. Understanding and leveraging this relationship allows traders and stakeholders to optimize their decision-making and profitability. In this project, we analyzed historical gasoline and crude oil pricing data to explore their cointegration characteristics thoroughly. We employed four distinct forecasting models to predict gasoline prices for the year 2024:

1. **ECM** (Error Correction Model)
2. **VAR** (Vector Autoregression):
3. **ARIMAX** (Auto-Regressive Integrated Moving Average with eXogenous Variables)
4. **LSTM** (Long Short-Term Memory):

Using these predictions, we developed tailored pricing strategies designed for effective trading. Subsequently, we evaluated the efficacy of each forecasting approach by calculating and comparing their respective profitability, providing valuable insights into model performance and the strategic implications for trading in energy commodities. We also hypothesize that LSTM will have the ability to capture the intricate collinear relationship between crude oil and gasoline prices most effectively.

2. Data Overview

This dataset has been pulled from the U.S. Energy Information Administration (eia) and contains 9 columns and 380 rows. This data spans over the years of 1993 to 2024. It contains information about Crude Oil and Gasoline prices, U.S, % utilization of refinery refinery operable capacity, Average Ending Stock for Motor Gasoline, Average Ending Stocks excluding SPR of Crude Oil, Product Supplied of Finished Motor Gasoline, Heating Degree Days, and Cooling Degree Days.

3. Exploratory Data Analysis

This dataset starts in 1993 and ends in 2024 with 380 total observations and 9 features.

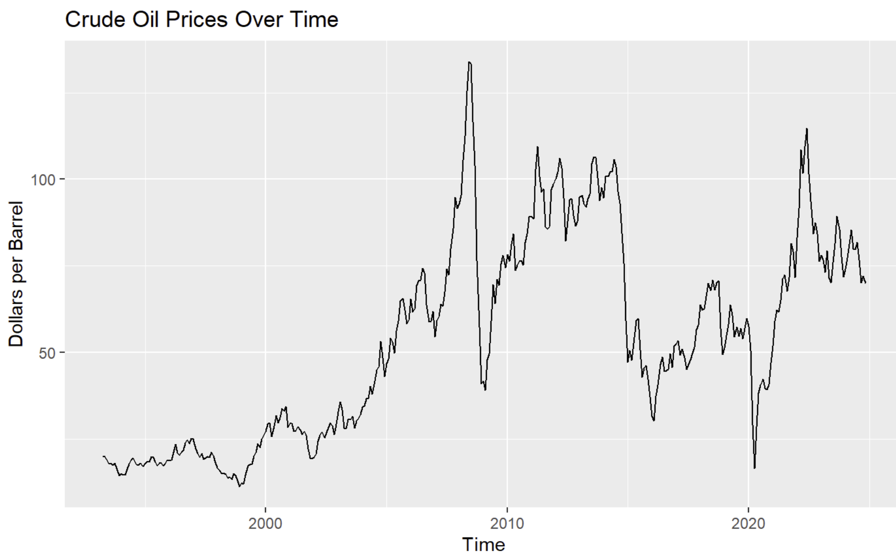


Fig 3.1: Crude oil Prices over Time

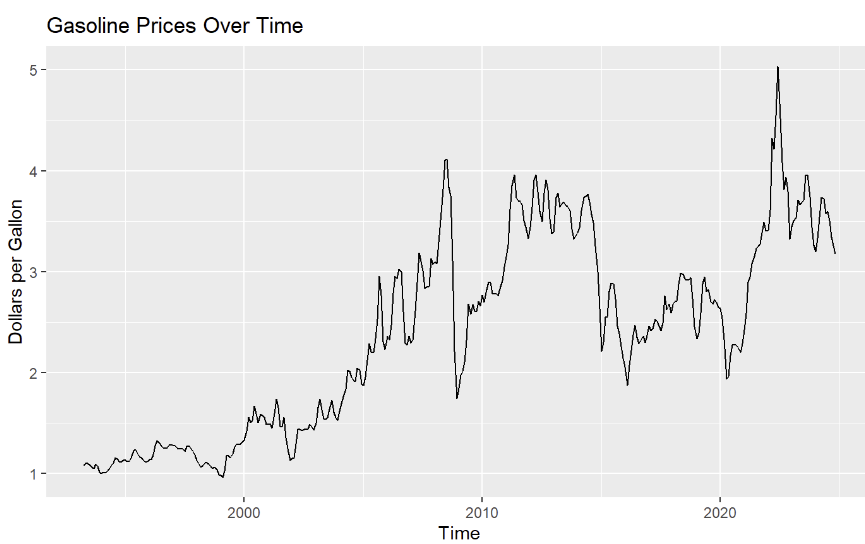


Fig 3.2: Gasoline Prices over Time

Observations from EDA of Prices over Time: The prices of both Crude oil and Gasoline generally increase over time, indicating a growing trend.

Cointegration test:

The spread (residuals from the cointegration regression) is stationary (p-value = 0.027 < 0.05).

Conclusion: Crude oil and gasoline prices are cointegrated!

This means they share a long-term equilibrium relationship, and deviations (the spread) are mean-reverting.

```
	Augmented Dickey-Fuller Test

data:  na.omit(spread)
Dickey-Fuller = -3.6599, Lag order = 7, p-value = 0.02739
alternative hypothesis: stationary
```

Stationarity tests:

Since stationarity is a key assumption for VAR and ARIMAX models, the ADF test was conducted, along with ACF and PACF plots, to assess stationarity and explore AR and MA terms relevant for ARIMAX modelling. Since the p-value was > 0.05 for the ADF tests, our time series are non-stationary. We differenced the time series by lag order 1 and achieved stationary with the ADF test resulting in a p-value < 0.05.

```
	Augmented Dickey-Fuller Test

data:  na.omit(crude_ts)
Dickey-Fuller = -2.5724, Lag order = 7, p-value = 0.3354
alternative hypothesis: stationary

	Augmented Dickey-Fuller Test

data:  na.omit(gasoline_ts)
Dickey-Fuller = -2.4815, Lag order = 7, p-value = 0.3738
alternative hypothesis: stationary
```

Fig 3.3: ADF Test Results for Crude Oil and Gasoline Prices

```
        Augmented Dickey-Fuller Test

data:  train_crude_diff
Dickey-Fuller = -7.2988, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary


        Augmented Dickey-Fuller Test

data:  train_gasoline_diff
Dickey-Fuller = -9.1872, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

Fig 3.4: ADF Test Results for Crude Oil and Gasoline Prices After Differencing

From the ACF plots for crude oil and gasoline prices, we see a slow decline before differencing, suggesting stationarity. However, after differencing, the ACF plots for both cuts off sharply, indicating that the series has been made stationary and is ready for differencing.

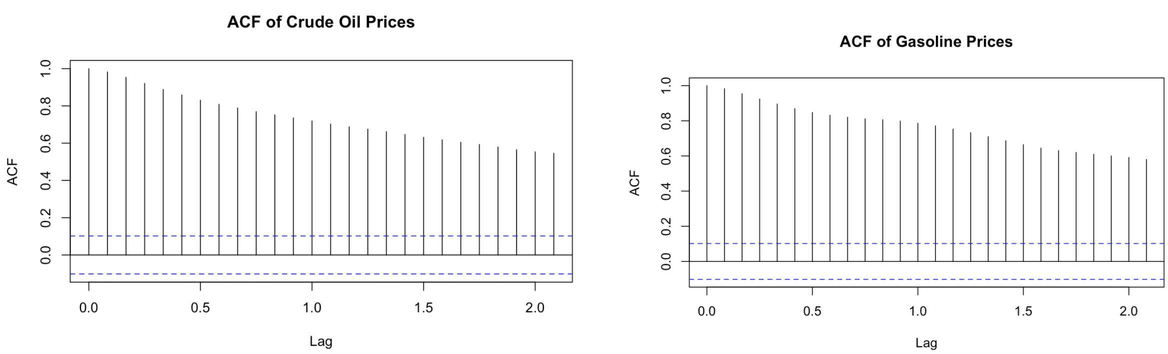


Fig 3.5: ACF of Crude Oil and Gasoline Prices Before Differencing

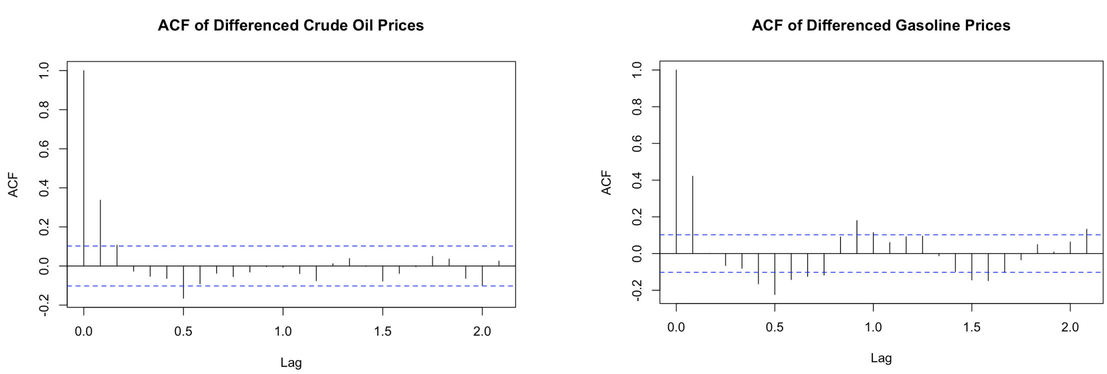


Fig 3.6: ACF of Crude Oil and Gasoline Prices After Differencing

For the ARIMAX model, it is crucial to ensure that both the dependent and exogenous variables are stationary. To assess stationarity, we performed the ADF test on the selected exogenous variables. A p-value below 0.05 indicates stationarity, while a higher p-value suggests the need for differencing.

The variables considered for stationarity testing were:

- U.S. Percent Utilization of Refinery Operable Capacity (p = 0.01)
- Heating Degree Days (p = 0.01)
- Cooling Degree Days (p = 0.01)
- Avg of U.S. Ending Stock of Motor Gasoline (p = 0.729)
- Average of U.S. Ending Stocks excluding SPR of Crude Oil (p = 0.34)
- U.S. Product Supplied of Finished Motor Gasoline (p = 0.24)

Among these, **Avg of U.S. Ending Stock of Motor Gasoline, Average of U.S. Ending Stocks excluding SPR of Crude Oil, and U.S. Product Supplied of Finished Motor Gasoline** had p-values above 0.05, indicating non-stationarity. To address this, first-order differencing was applied, after which their p-values dropped to 0.01, confirming stationarity.

4. Modeling

4.1 ECM Analysis

ECM (Error Correction Model) forecast cointegrated time series, capturing both short-term fluctuations and long-term equilibrium relationships.

A. Model Fitting

We first performed cointegration regression to estimate the long-term equilibrium relationship between gasoline and crude oil prices. Subsequently, we derived lagged error correction terms and fitted the ECM model using changes in gasoline prices, crude oil prices, and the lagged spread.

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.49888 -0.04899 -0.01350  0.05216  0.42383 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.078521   0.015240   5.152 4.22e-07 ***
delta_crude  0.020909   0.001077  19.422  < 2e-16 ***
spread_lag1 -0.110270   0.020708  -5.325 1.77e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1065 on 365 degrees of freedom
Multiple R-squared:  0.5326,	Adjusted R-squared:  0.5301 
F-statistic:   208 on 2 and 365 DF,  p-value: < 2.2e-16
```

Fig. 4.1.1 ECM Model Summary

B. Model Evaluation

The ECM model effectively captures the general price movement of gasoline but exhibits deviations at certain points. The predicted price follows a mean-reverting expectation, suggesting potential arbitrage opportunities. When the forecasted price significantly differs from the actual price, a trading opportunity arises. Specifically, if the forecasted price is higher than the actual price, buying gasoline contracts may be advantageous, whereas if the forecasted price is lower, selling gasoline contracts could be a profitable strategy.

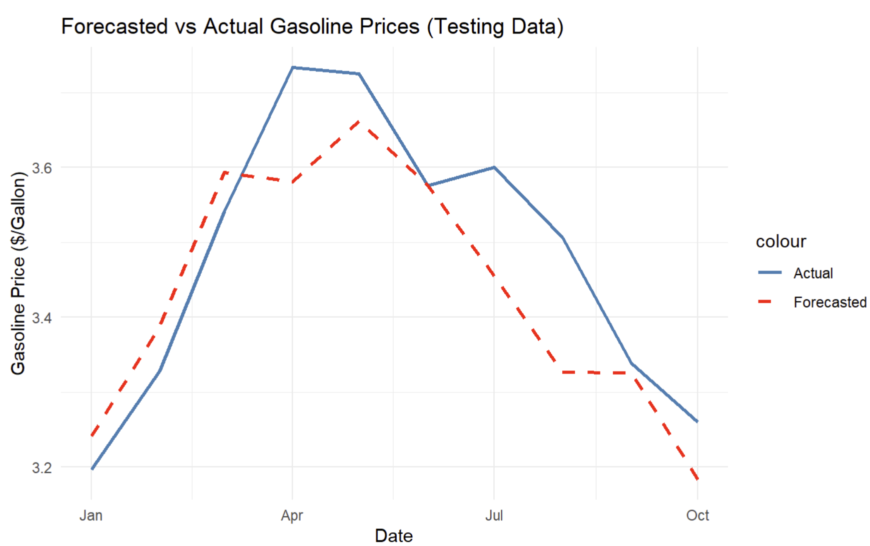


Fig 4.1.2: Forecasted vs Actual Gasoline Price

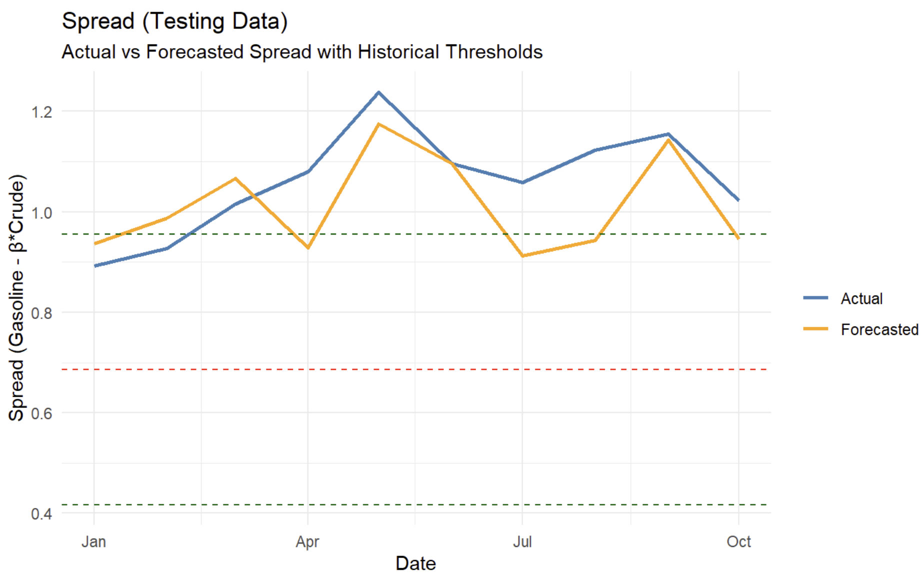


Fig 4.1.3: Forecasted vs Actual Spread

C. RMSE and MAPE Metrics

We calculate the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) using the accuracy function to evaluate the model's performance on the training set.

Metric	Value
RMSE ($/gallon)	$0.0975
MAPE (%)	2.2400

Table 4.1.4: ECM Performance Metrics

Evaluation Summary: **RMSE = $0.0975 per gallon** indicates the average deviation of the predicted gasoline prices from actual prices. A lower RMSE suggests a well-fitted model with minimal prediction errors.

MAPE = 2.24% suggests that, on average, the model's predictions deviate by only 2.24% from actual values, indicating a high level of accuracy in forecasting gasoline prices.

4.2 VAR Analysis

VAR (Vector Autoregression) captures the interdependencies among multiple time series variables by modeling each variable as a function of past values of itself and other variables.

A. Model Fitting

We determined the optimal number of lags for the VAR model using the Akaike Information Criterion (AIC), which indicated an optimal lag length of 10. The VAR model was then fitted using the selected optimal lag to effectively capture the dynamic interdependencies between changes in gasoline and crude oil prices.

```
Residual standard error: 0.1202 on 337 degrees of freedom
Multiple R-Squared: 0.4504,  Adjusted R-squared: 0.4178
F-statistic: 13.81 on 20 and 337 DF,  p-value: < 2.2e-16



Covariance matrix of residuals:
                    train_crude_diff train_gasoline_diff
train_crude_diff             23.4307             0.38876
train_gasoline_diff           0.3888             0.01444

Correlation matrix of residuals:
                    train_crude_diff train_gasoline_diff
train_crude_diff              1.0000              0.6684
train_gasoline_diff           0.6684              1.0000
```

Fig. 4.2.1 VAR Model Summary

B. Model Fitting

The VAR model captures the broader trends in gasoline and crude oil prices, making it a useful tool for understanding long-term price movements. However, its ability to react to rapid fluctuations is limited, potentially causing lags in signal generation. In gasoline prices, if the forecast struggles to align with sharp peaks and troughs, traders relying on it for short-term decision-making might face delayed entries or exits, increasing exposure to price swings.

For crude oil, while the forecast tracks the general 60–90 price range, any smoothing of fluctuations or lagging responses could result in missed opportunities in volatile markets. Compared to more adaptive models, VAR's reliance on historical lag structures may hinder its effectiveness in fast-paced trading environments. Still, it remains valuable for strategic decision-making, helping traders anticipate longer-term trends and identify macro-level price movements.

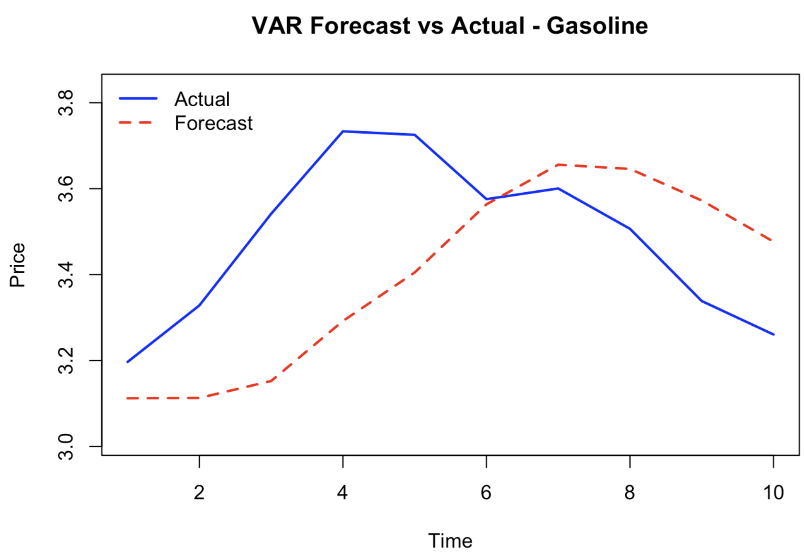


Fig 4.2.2: Forecasted vs Actual Gasoline Price

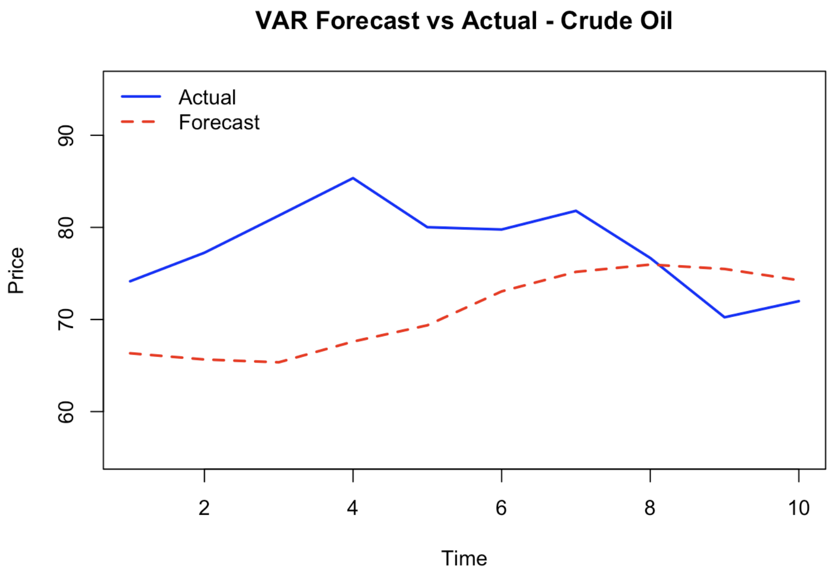


Fig 4.2.3: Forecasted vs Actual Crude Oil Price

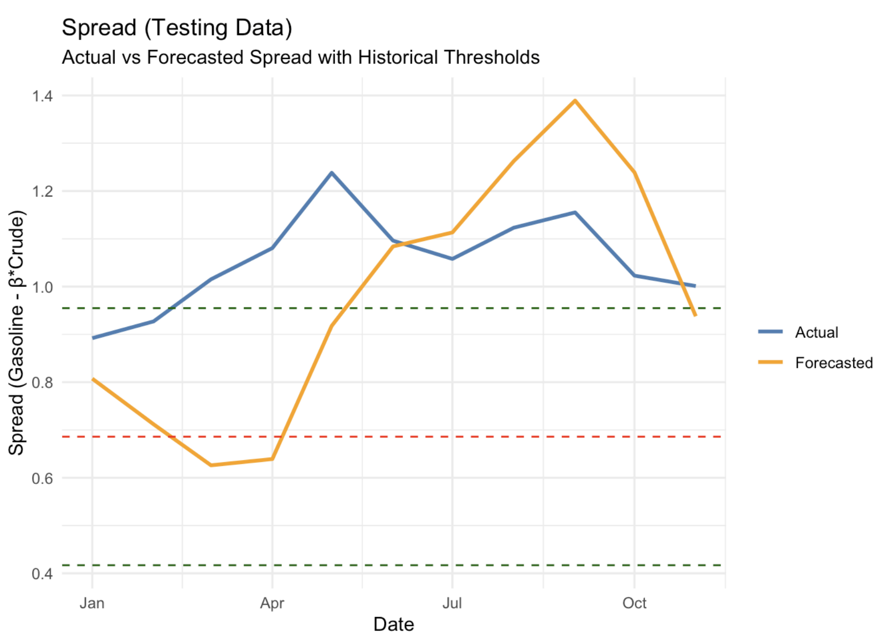


Fig 4.2.4: Forecasted vs Actual Spread

C. RMSE and MAPE Metrics

The performance metrics for the VAR forecasting model indicate robust predictive capabilities for both gasoline and crude oil prices. For gasoline, the Root Mean Square Error (RMSE) is $0.239 per gallon, and the Mean Absolute Percentage Error (MAPE) is

9.59%, suggesting good accuracy and reliability of the VAR model in forecasting gasoline price fluctuations. However, the model shows relatively higher prediction errors for crude oil, with an RMSE of $5.64 per gallon and a MAPE of 10.23%, indicating slightly less precision when predicting crude oil price variations.

Metric	Value
RMSE ($/gallon)	$0.239
MAPE (%)	9.59

Table 4.2.5: VAR Gasoline Performance Metrics

Metric	Value
RMSE ($/gallon)	$5.64
MAPE (%)	10.23

Table 4.2.6: VAR Crude Oil Performance Metrics

These differences highlight the VAR model's stronger performance in forecasting gasoline prices compared to crude oil, potentially reflecting greater volatility or structural complexity within crude oil market data.

4.3 ARIMAX Analysis

ARIMAX (Auto-Regressive Integrated Moving Average with eXogenous variables) extends ARIMA by incorporating external variables that influence the dependent variable. It combines autoregression, differencing for stationarity, and moving averages while leveraging exogenous predictors to improve forecast accuracy.

A. Model Fitting

We fitted an ARIMAX model using auto.arima() with exogenous variables to capture external influences on gasoline and crude oil prices. The function selected **ARIMA(0,0,0)**, suggesting that the differenced series did not exhibit significant autoregressive or moving average structures, and the model primarily relied on the exogenous variables to explain variations in prices.

```
Series: train_gasoline_diff
Regression with ARIMA(0,0,0) errors

Coefficients:
      U.S. Percent Utilization of Refinery Operable Capacity %  Avg of US Ending Stock of Motor Gasoline
                                                        0.0019                                     0e+00
s.e.                                                    0.0027                                     1e-04
      Average of U.S. Ending Stocks excluding SPR of Crude Oil  U.S. Product Supplied of Finished Motor Gasoline
                                                         0e+00                                             0e+00
s.e.                                                     1e-04                                             1e-04
      Heating Degree Days  Cooling Degree Days
                   -3e-04               -1e-04
s.e.                1e-04                1e-04
```

Figure 4.3.1: ARIMAX Model Summary for Gasoline Prices

```
Series: train_crude_diff
Regression with ARIMA(0,0,0) errors

Coefficients:
      U.S. Percent Utilization of Refinery Operable Capacity %  Avg of US Ending Stock of Motor Gasoline
                                                        0.1119                                    -1e-04
s.e.                                                    0.0949                                     1e-04
      Average of U.S. Ending Stocks excluding SPR of Crude Oil  U.S. Product Supplied of Finished Motor Gasoline
                                                         0e+00                                             0e+00
s.e.                                                     1e-04                                             1e-04
      Heating Degree Days  Cooling Degree Days
                  -0.0057              -0.0025
s.e.               0.0018               0.0041
```

Figure 4.3.2: ARIMAX Model Summary for Crude Oil Prices

B. Model Evaluation

The ARIMAX forecast captures some price fluctuations, reflecting the influence of exogenous variables, but it lags behind actual price changes, reacting slowly to sudden spikes or drops. While the model could identify broader trends, its delayed responsiveness limits its ability to capture short-term volatility. This can impact trading by generating lagged signals, potentially leading to missed opportunities or increased risk in fast-moving markets.

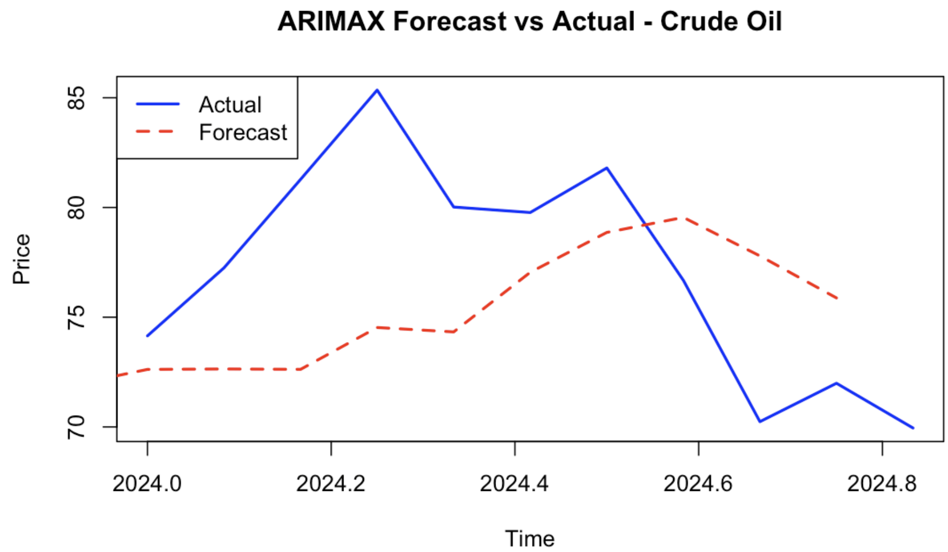


Fig 4.3.3: Forecasted vs Actual Gasoline Price

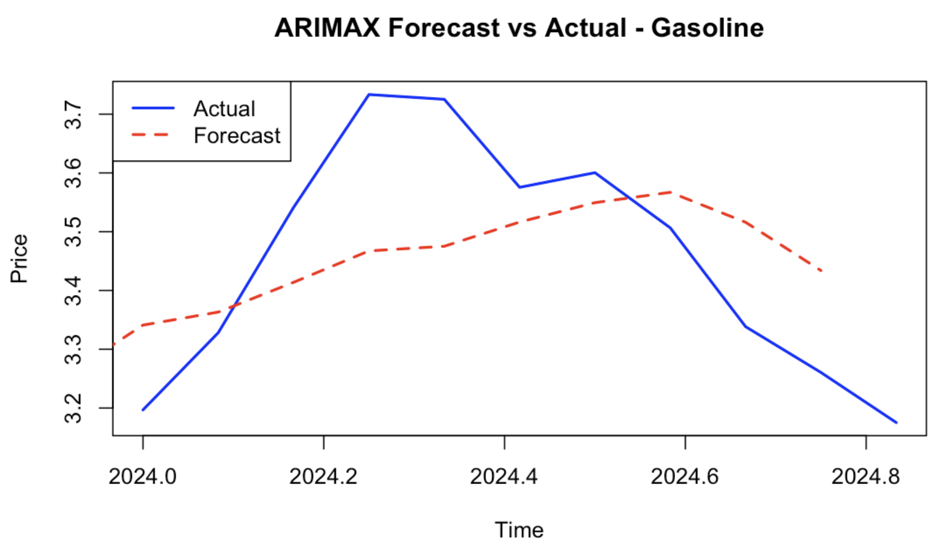


Fig 4.3.4: Forecasted vs Actual Gasoline Price

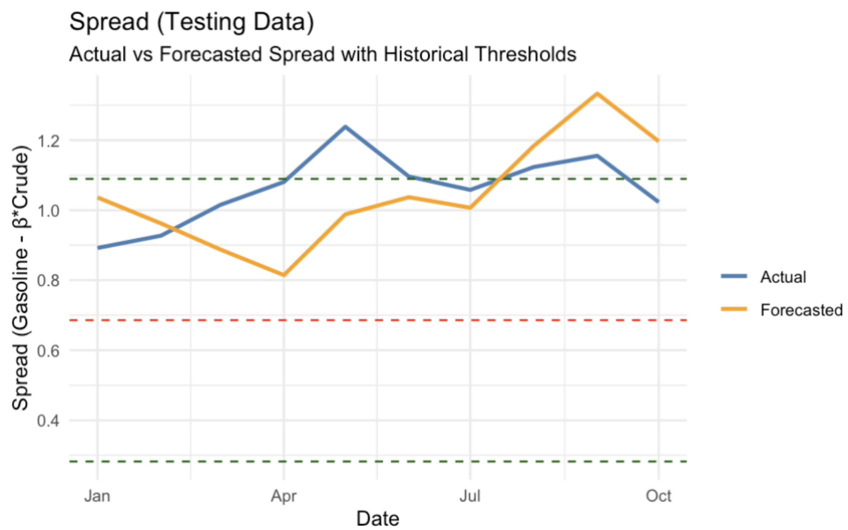


Fig 4.3.5: Forecasted vs Actual Spread

C. RMSE and MAPE Metrics

The ARIMAX model demonstrates strong predictive accuracy for gasoline prices, with a low RMSE of $0.13 and a MAPE of 3.84%. This aligns with its ability to capture broader trends while still reacting to some price fluctuations. However, the relatively low error does not negate its lag in responding to sudden spikes or drops, which can impact short-term trading decisions.

Metric	Value
RMSE ($/gallon)	$0.13
MAPE (%)	3.84

Table 4.3.6: ARIMAX Gasoline Performance Metrics

Metric	Value
RMSE ($/gallon)	$5.13
MAPE (%)	6.5

Table 4.3.7: ARIMAX Crude Oil Performance Metrics

For crude oil, the higher RMSE of $5.13 and MAPE of 6.5% indicate greater difficulty in precisely tracking price movements. While the model effectively follows overall trends, its responsiveness to market shifts remains limited. This reinforces its strength as a tool for long-term strategic insights but highlights challenges in capturing rapid price changes crucial for short-term trading strategies.

4.4 LSTM Analysis

LSTM (Long Short-Term Memory) recurrent neural network designed to learn long-term dependencies in sequential data effectively.

A. Model Fitting

We compiled the LSTM with the Adam optimizer at a suitable learning rate (0.001) and used Mean Squared Error (MSE) as our loss function to measure prediction error on the scaled Gasoline prices. We then trained the model for a defined number of epochs (50) with a chosen batch size (16), reserving a fraction of the training data (10%) for validation to monitor and mitigate overfitting as the network's weights updated.

B. Model Evaluation:

The LSTM model appears to best forecast the price of gasoline for the given time period. You can see that while the LSTM model generally captures the upward and downward trends in gasoline prices, it sometimes under or overestimates the magnitude of certain peaks and troughs which is common in real-world forecasting where the model's outputs will not perfectly match observed data, but should still track the broader directional movements over time.

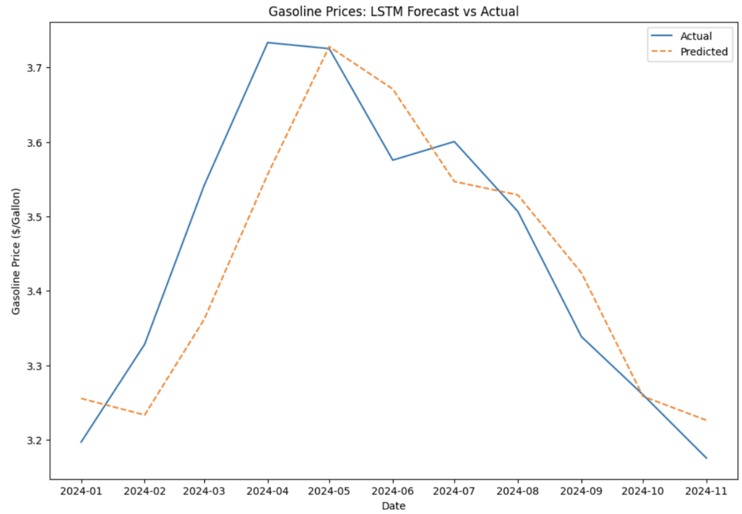


Fig 4.4.1 - Predicted vs Actual Gasoline Prices

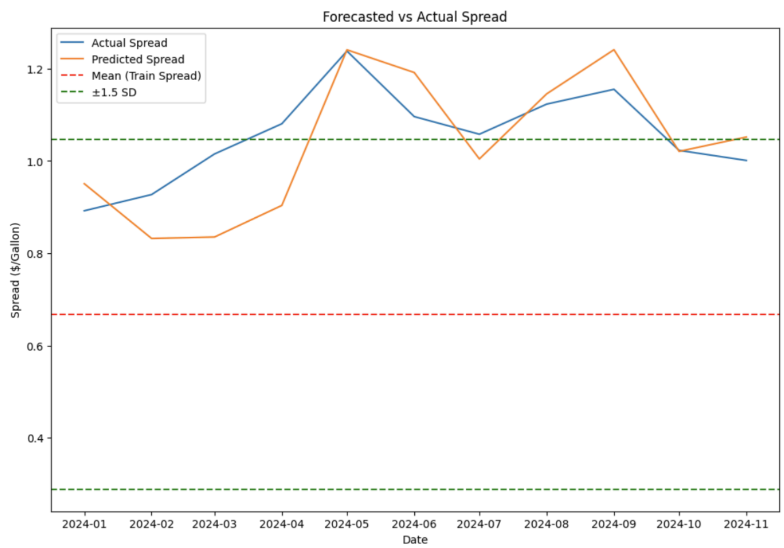


Fig 4.4.2 - Predicted vs Actual Spread

C. RMSE and MAPE Metrics

The root-mean-squared error of $0.0947 suggests that, on average, the model's predictions deviate from the actual gasoline price by about 9.47 cents per gallon. A mean absolute percentage error of 2.15% indicates the model's forecasts are, on average, about 2.15% away from the actual observed prices—a relatively strong performance for this type of time-series prediction.

Metric	Value
RMSE ($/gallon)	$0.0947
MAPE (%)	2.15%

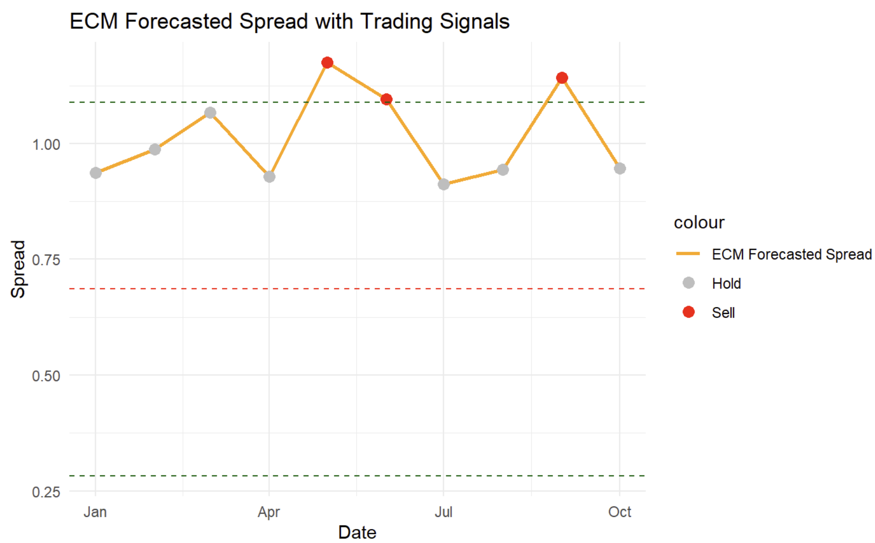
Table 4.4.3 - Accuracy Metrics LSTM

5. Trading Strategy

The trading strategy for crude oil and gasoline spread is designed using four different forecasting models: ECM, LSTM, VAR, and ARIMAX. Each model predicts the expected spread between crude oil and gasoline prices, generating trading signals based on deviations from the historical mean. Across all models, the strategy follows a buy-low, sell-high approach, entering trades when spreads deviate significantly and profiting from their return to equilibrium. The effectiveness of each model is compared based on profitability.

A. ECM

The first graph represents the ECM forecasted spread, where sell signals appear when the spread reaches high levels, indicating potential overpricing in the market. The second graph maps these sell signals onto the actual gasoline price, allowing us to gauge the profitability of the trades. The alignment between high spread values and corresponding gasoline price peaks suggests that the ECM model successfully identifies mean-reversion opportunities, enabling well-timed trades.



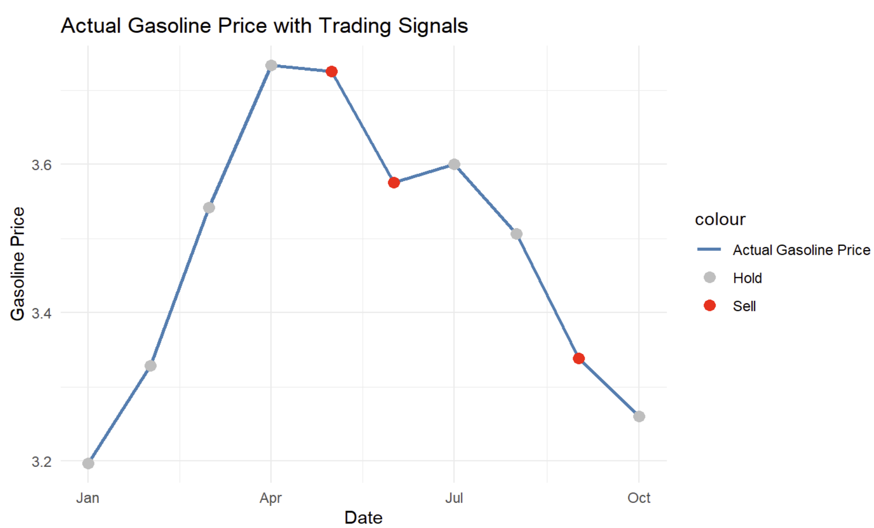


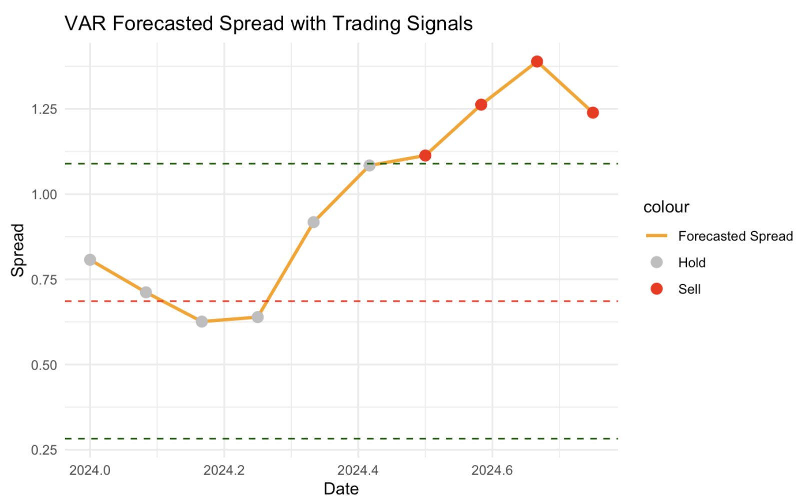
Fig 5.1 ECM Trade Signals

The cumulative profit graph demonstrates the effectiveness of the ECM-based trading strategy, with a final total profit of $202.75. The stepwise nature of the profit curve indicates that trades were executed selectively, with major gains occurring at key points of mean reversion. The sharp increase in profit around April suggests that the model successfully identified profitable spread deviations, while the stability in mid-year reflects controlled risk and minimal drawdowns. By late 2024, profits steadily increased, reinforcing ECM's ability to capture mean-reverting opportunities in the crude oil-gasoline spread.

Fig 5.2 Cumulative Profit from ECM trading strategy

B. VAR

This visual representation underscores the effectiveness of the trading strategy, highlighting points where selling would have optimized profitability by capturing peaks before subsequent declines. We see that there are 6 hold and 4 sell signals which generates a cumulative profit of $425.



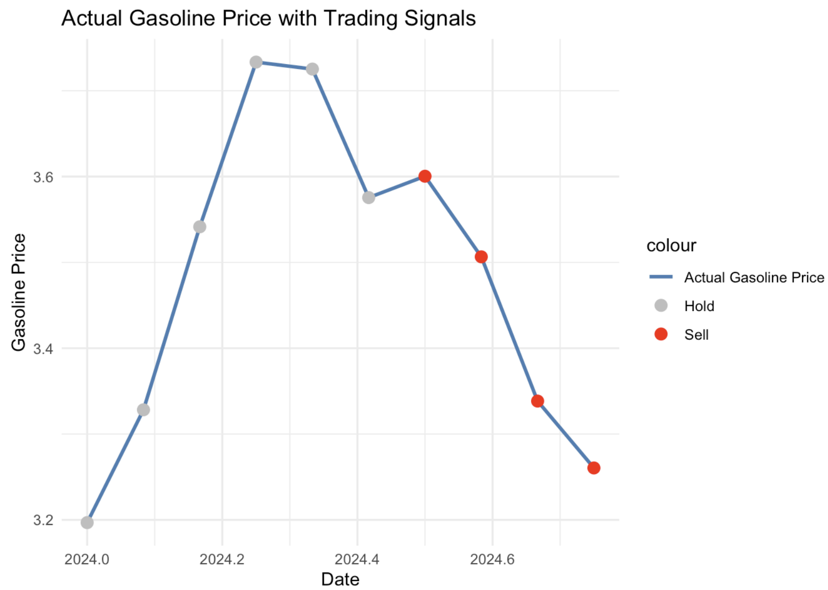


Fig 5.3 VAR Trade Signals

The graph illustrates the cumulative profit obtained using the VAR-based spread trading strategy in 2024. The strategy generated substantial profit growth starting mid-year, indicating its effectiveness in identifying favorable trading opportunities during that period.

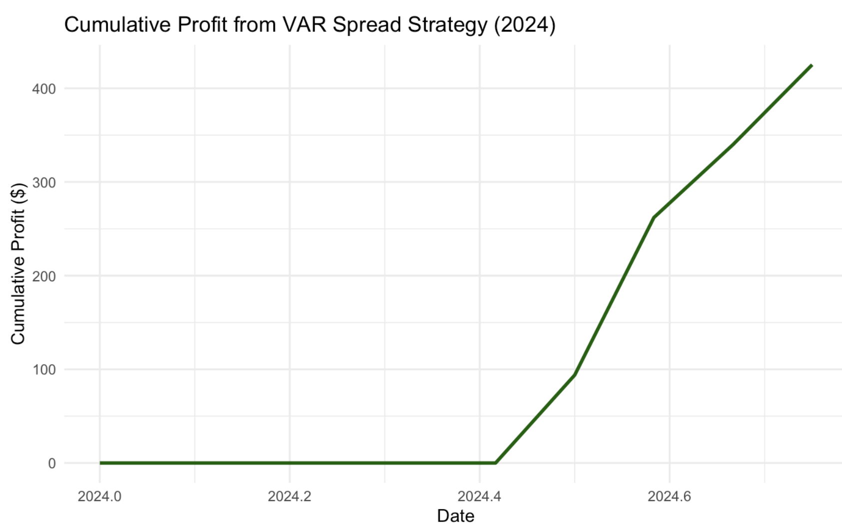


Fig 5.4 Cumulative Profit from VAR trading strategy

C. ARIMAX

The graphs depict the performance of the ARIMAX model in forecasting the spread between gasoline and crude oil prices, along with associated trading signals. It demonstrates how the model successfully identified peaks in gasoline prices, providing profitable selling opportunities before subsequent price declines. We see the model generates 7 hold signals and 3 sell signals making a total profit of $246.

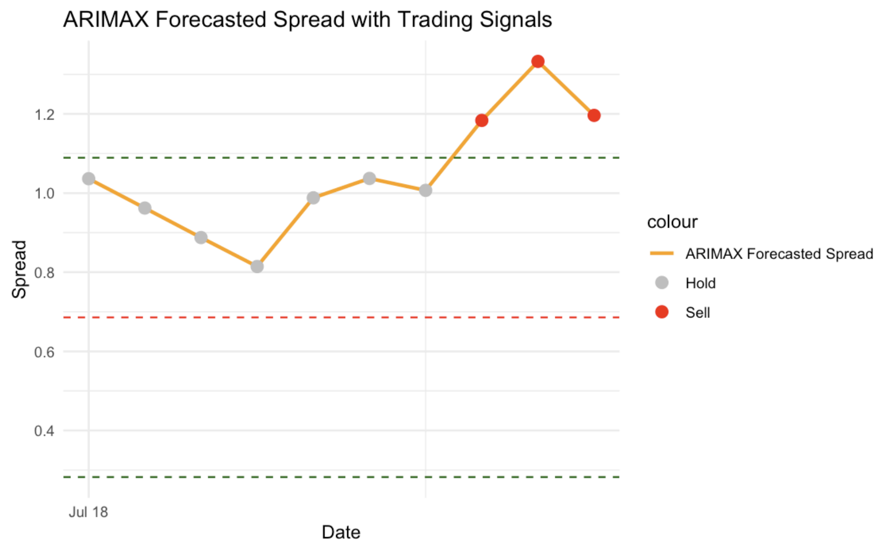


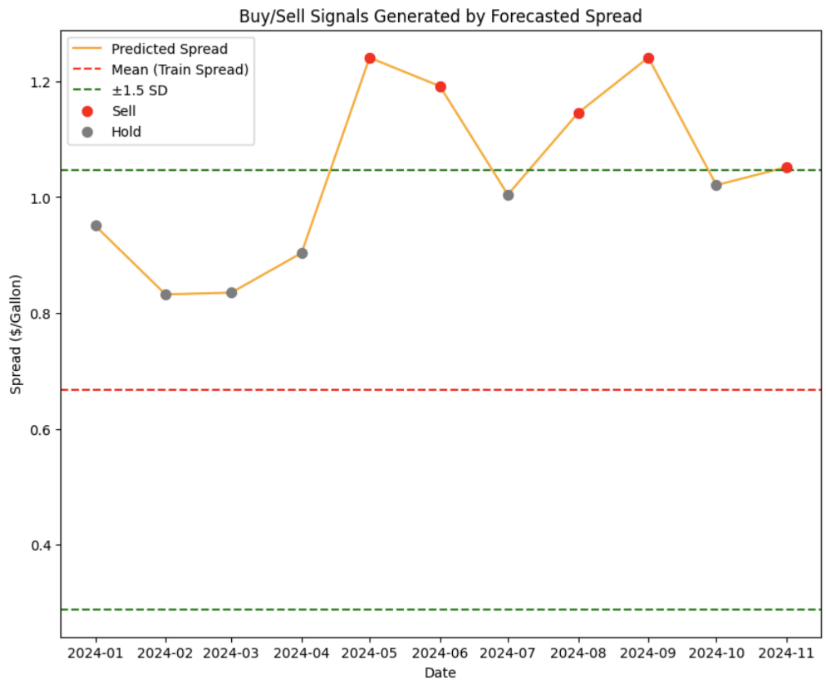
Fig 5.5 ARIMAX Trade Signals

The ARIMAX forecast captures price fluctuations influenced by exogenous variables but lags behind actual price changes, reacting slowly to sudden spikes or drops. This delayed responsiveness likely contributed to early losses due to misaligned trade signals. However, as market conditions realigned with historical patterns, the model effectively identified broader trends, leading to a strong upward profit trend in the second half, surpassing breakeven and reaching $246. While the model can generate profitable signals over time, its lag in capturing short-term volatility may result in missed opportunities or increased risk in fast-moving markets.

Fig 5.6 Cumulative Profit from ARIMAX trading strategy

D. LSTM

The two graphs below illustrate the trade signals generated by the spread forecasted by the LSTM model. While the forecasted spread models the actual spread very closely, we can see that does not always translate into an optimal trade on the gasoline price. At times, even though the forecasted spread may indicate a sell, gasoline may tend to move in the opposite direction. However, this allows us to gain an accurate understanding of the effect that spread analysis has on trading gasoline.



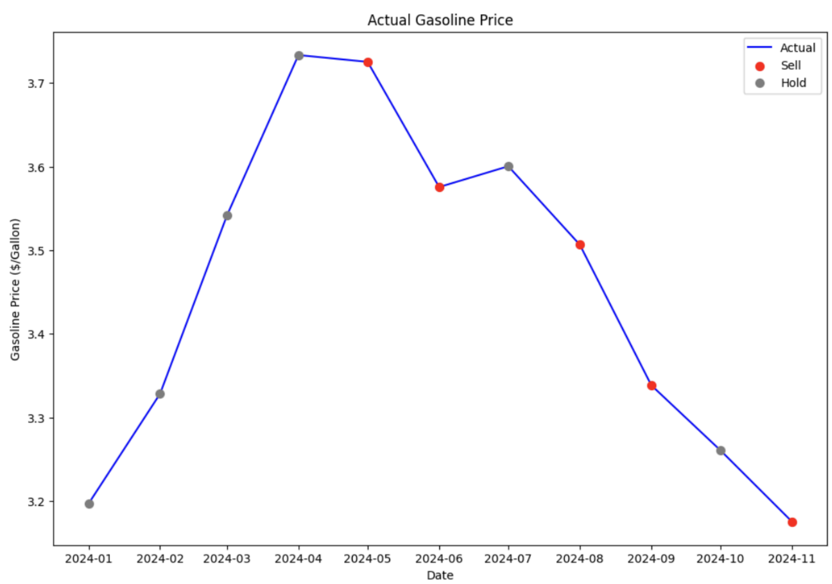


Fig 5.7 LSTM Trade Signals

The graph below shows the cumulative trading profit over the 2024 year. The trading strategy keeps us from losing money during the first half of the year and provides trade signals with accuracy of 75% throughout the test period. Our LSTM trading model makes a cumulative profit of $370.85.

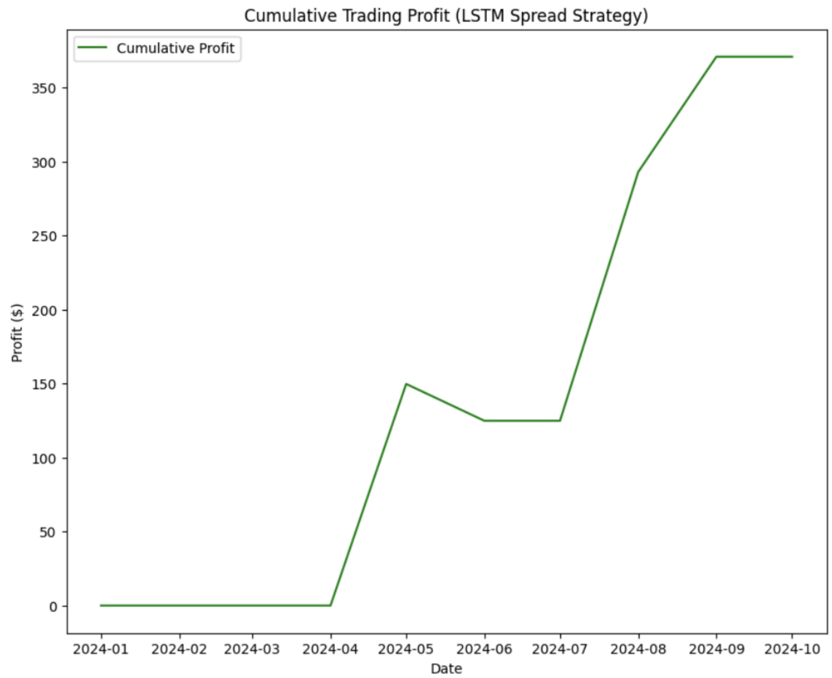


Fig 5.8 Cumulative Profit From LSTM Trading Strategy

6. Model Comparison and Discussion

Model	Number of Hold Signals	Number of Sell Signals	Cumulative Profits
ECM	7	3	$202.75
VAR	6	4	$425.15
ARIMAX	7	3	$246.00
LSTM	6	4	$370.85

Fig 6.1 ECM, VAR, ARIMAX, LSTM Model Comparison

In our analysis, the VAR model produced the highest overall cumulative profit at $425.15, followed by LSTM at $370.85. ARIMAX and ECM delivered lower returns ($246.00 and $202.75, respectively) but still generated actionable signals. While higher profitability typically indicates stronger performance, other factors—such as risk tolerance, forecast accuracy, and the frequency of signals—must also be weighed when choosing a forecasting model.

7. Conclusion

In this report, we evaluated the performance of four predictive models—ECM, VAR, ARIMAX, and LSTM—in generating hold and sell signals and assessing their profitability. Each model exhibited distinct strengths and weaknesses. VAR, a traditional econometric model, demonstrated strong predictive capabilities in this context, while LSTM, a deep learning-based approach, showed promise in capturing complex temporal dependencies. ARIMAX, incorporating external regressors, provided moderate profitability, and ECM, designed to model long-term relationships, yielded the lowest cumulative returns.

Ultimately, while VAR and LSTM emerged as the most profitable models in this case, the selection of an optimal forecasting approach depends on various factors, including data characteristics, risk tolerance, and model interpretability. Future work could explore hybrid approaches or alternative hyperparameter tuning to further refine predictive accuracy and profitability.