Quant Commodities Task by Alpha Alternatives

Gaurav Sangwan

September 23, 2024

1 Introduction

This report is accompanied by the Task given for developing a trading strategy for natural gas futures, using two independent approaches, Statistical Arbitrage and Machine Learning Techniques. The entire analysis is based on the dataset provided for NG1 and NG2, natural gas inventory data, market participant positioning data, and weather data (HDD and CDD). The Statistical Arbitrage approach identifies mean-reverting patterns between NG1 and NG2 prices and Time series and Crosssectional Momentums, while the machine learning approach uses feature engineering to create a training and testing data with all the given data including the prices to predict the future spread between the two and place trade based on rules associated with it. This report compares the peformance of these two approaches and provides conclusions and recommendataions for future work.

2 Approach 1: Statistical Arbitrage

I worked on the hypothesis the Natural gas and its future can be subjected to pairs trading, to verify the same I used **co-integration**, a statistical method that identifies pairs of assets that maintain a stable long-terrm relationship by finding such pairs, we can bettter predict when prices are likely to converge or spread.

The process to verify that NG1 and NG2 have a strong relationship for trading, I used the given data and performed Ordinary least squares(OLS) regression between the log prices of both NG1 and NG2, to calculate the residual. Then I did, **Stationarity test** by applying Augmented Dickey Fuller (ADF) test to the residuals ϵ_t. A p-value of ¡ 0.05 means that results are stationary, indicating that the price difference is mean-reverting. And if the pair have most negative ADF test statistic then it indicates a strong mean reverting relationship.

For our given data, I obtained the following results. Therefore, I have a strong mean reverting relationship between NG1 and NG2

2.1 Mean Reversion

Now, since we know that NG1 and NG2 have a strong mean-reverting relationship, to generate the signals for buying and selling NG2 based on NG1 we would need to calculate z scores. For that we use the following approach:

- Calculate residuals for each timestamp using the OLS formula.

Pair	p value	test_statistic
NG1 & NG2	0.000035	-4.899109

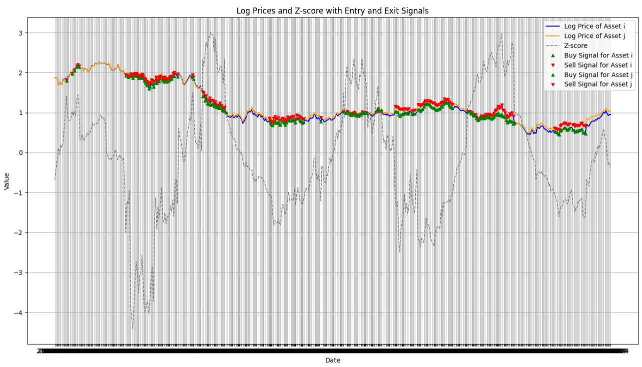

Figure 1: Reversal Signals

- Calculate the rolling β using the covariance and variance of both NG1 and NG2 prices at each timestamp over a 90-day rolling window.
- The rolling alpha for each timestamp will be the as per the following formula $\alpha_t = \mu_{\log(p_{ng2,t})} - \beta_{\log(p_{ng1,t})}$.
- Finally the z-score will be calculated for each time stamp using the following formula $z_t = (\epsilon_t - \mu_t)/\sigma_t$

  The reason for calculating the z score is to get the direction of the idea correctly while avoiding outliers. Meanwhile I also tried to use returns rather than prices or log prices to avoid trending data but it gave almost same result.

  So, Now we will have following Trading signals

  Entry Signals :

  - Short 1 unit of NG2 and Long β_t units of NG1 if $z_t > 1$.
  - Long 1 unit of NG2 and Short β_t units of NG1 if $z_t < -1$.
  - This way we maintain dollar neutrality by balancing the value of long and short positions to minimize market exposure.

  Exit Signals : Close the positions when the z score reaches any of the following thresholds : $z_t >= -threshold$ or $z_t <= threshold$. For determining threshold I tested with 0, 0.1,0.2, 0.3, 0.4 but got the best Sharpe ratio of 1.353659 at threshold at 0.

  Final results for the above strategy shows the following plots for turnover and signals. Also, the max drawdown is -10.610 and max drawdown duration is 107, also accompanied by the graphs.

Results for the strategy in a summarized way are as follows:

2.2 Momentum (Both TS and XS)

For momentum based approach in TS I used the following strategy:

- Calculate the short-term and long-term moving averages of returns.

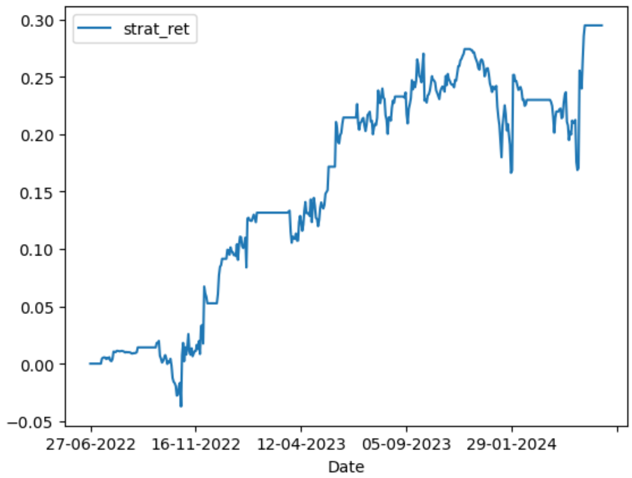

Figure 2: Reversal Strategy Returns

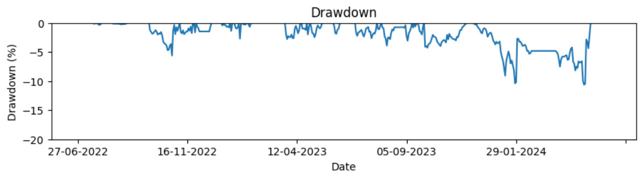

Figure 3: Reversal Drawdown

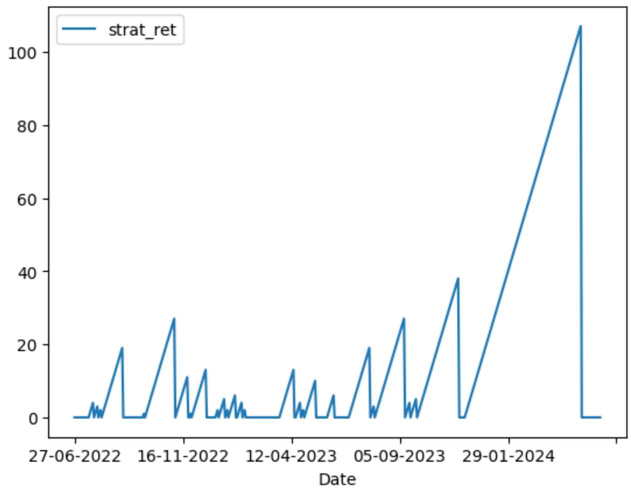

Figure 4: Reversal Drawdown duration

	strategy
Return (Ann.)	15.29%
Volatility (Ann.)	15.34%
Sharpe Ratio	1.00
Hit Rate	35.39%
Max Drawdown	-10.61%
Max Duration	107.00 days
Holding Period	59.96 days

Figure 5: Reversal results

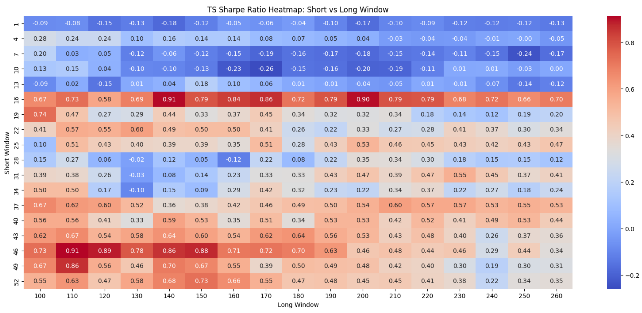

Figure 6: TS Sharpe Ratio heatmap

- Compute the z-score of the past 10-day returns, which measures the deviation of the short-term average from the long-term average.
- Apply a tanh function to curtail the z-scores and produce a portfolio signal.

For Time series the SR heat map looks as follows :
Strategy Return for TS Momentum strategy is as follows
Max drawdown is -0.809 and the drawdown is as follows
For XS I used the following:

- Calculate the short-term and long-term moving averages of returns.

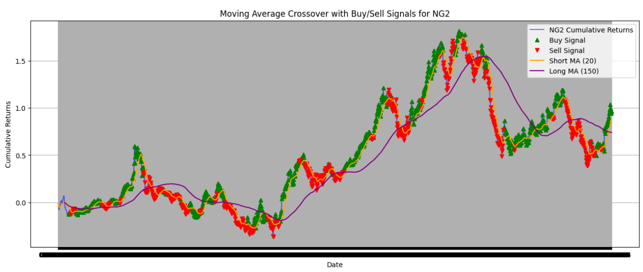

Figure 7: TS MA Crossover

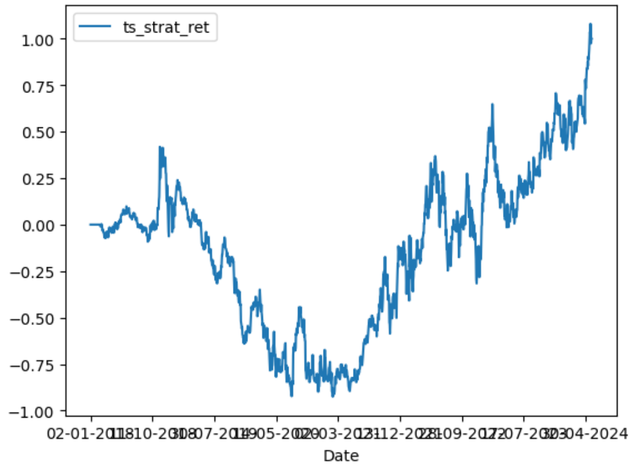

Figure 8: TS Strategy Return

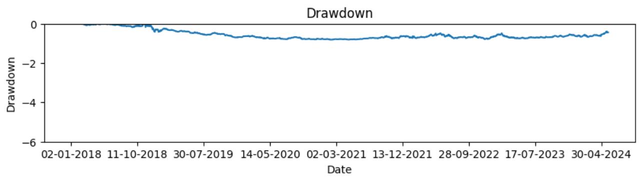

Figure 9: TS Drawdown

- Compute the difference between the short-term and long-term averages.
- Rank the differences across assets at each time point.
- Demean the ranked signal to achieve dollar neutrality.
- Normalize the portfolio to be fully invested (sum of portfolio weights = 1).

Returns for XS Strategy is as follows Max drawdown is -0.469 with drawdowns as follows

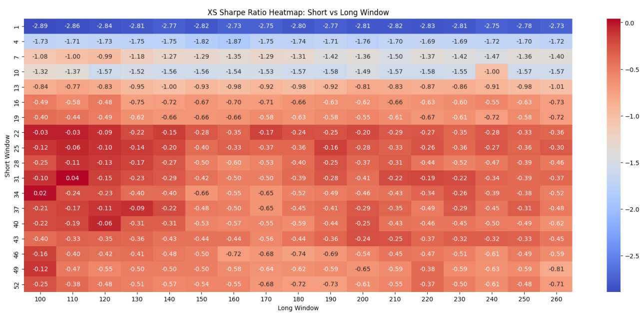

Figure 10: XS Sharpe Ratio Heatmap

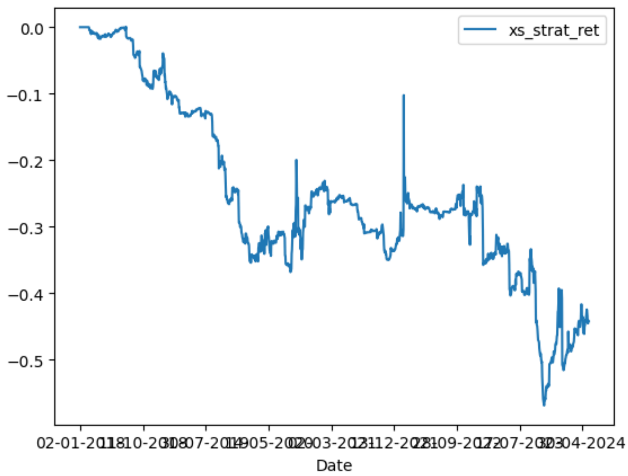

Figure 11: XS Strat Returns

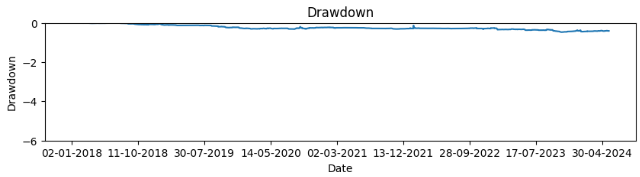

Figure 12: XS Drawdown

3 Approach 2: Machine Learning Based Approach

The Idea behind the machine learning based approach was that we combine other given data with mean reversion principles to identify the trading oppurtunities in Natural Gas futures.

3.1 Feature Engineering

- Calculate the difference between NG1 and NG2 prices (spread).
- Calcualte the moving average of NG1 and NG2 prices both long and short.
- Calcualte the cummulative sum of heating and cooling degree days.
- Calculate the inventory levels and their long and short moving averages.
- Calculate the market participants positioning metrics mainly long-short ratios for producers, swaps, managed money, other reportable and non-reportables.

This feature engineering results in a dataframe of 3971 rows and 14 columns.

3.2 Model Training

Train a random forest regressor model on the engineered features to predict the spread between NG1 and NG2 prices. Then utilising a walk-forward optimization approach to tune the model's hyperparameters.

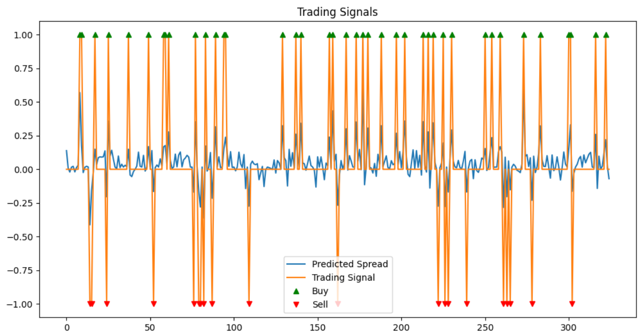

Figure 13: ML Trading Signals

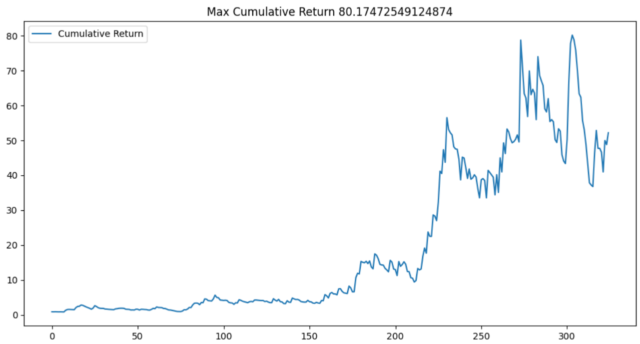

Figure 14: ML Cumulative return

3.3 Trading Rules

- If the predicted spread is above a certain threshold (**1.16 standard deviation**) go long NG1 and short NG2.
- If the predicted spread is above a certain threshold (**-1.16 standard deviation**) go short NG1 and long NG2.

Chosen threshold for spread is standard deviation of 1.16 and -1.16.

3.4 Results

Obtain Sharpe Ratio of 9.05 which was extremely high. Trading signals can be seen in Figure 13. Cummulative return and its plot is in Figure 14.

4 Comparision and Conclusion

The Statistical Arbitrage approach identified a strong mean-reverting relationship between NG1 and NG2 prices, and the trading strategy based on this approach generated a Sharpe Ratio of 1.0000. The strategy also exhibited a maximum drawdown of -10.610 and a maximum drawdown duration of 107 days.

In addition to the mean-reversion strategy, we also explored momentum-based approaches, including Time Series (TS) and Cross-Sectional (XS) momentum. The TS momentum strategy generated a Sharpe Ratio of 1.353659, while the XS momentum strategy generated a negative Sharpe Ratio.

The Machine Learning approach, which combined feature engineering with a random forest regressor model, generated a Sharpe Ratio of 9.05, which is significantly higher than the Sharpe Ratios obtained from the Statistical Arbitrage approach. The trading strategy based on this approach also exhibited a higher cumulative return.

In conclusion, both the Statistical Arbitrage and Machine Learning approaches have their strengths and weaknesses. The Statistical Arbitrage approach identified a strong mean-reverting relationship between NG1 and NG2 prices and generated a reasonable Sharpe Ratio. The Machine Learning approach, on the other hand, generated a significantly higher Sharpe Ratio and cumulative return.

Based on the results obtained, I would suggest using a more risk adjusted variant of ML approach in sync with Statistical Approach to obtain a more robust trading strategy.