

Pairs Trading Strategy on Gasoline and Heating Oil Futures

Applying Time Series Concepts to the Energy Markets

Group — Ramiro Díez-Liébana Orejas, Dominique Peytrignet

November 2025

Contents

1	Project Overview	2
1.1	Data Acquisition	2
1.2	Preprocessing	2
1.3	Time Series Modelling	2
1.4	Findings	4
1.5	Conclusion	5

1 Project Overview

This project aims to implement a pairs trading strategy on gasoline and heating oil futures contracts traded on the New York Mercantile Exchange (NYMEX). By leveraging time series analysis techniques, we seek to identify profitable trading opportunities based on the historical relationship between these two commodities. Both are refined from the same crude oil in shared refineries, face identical supply shocks, refining constraints, and crack-spread dynamics, and are influenced by overlapping seasonal demand patterns. These fundamental links create a strong, persistent correlation, making temporary divergences in their price ratio attractive for mean-reversion-based trading.

1.1 Data Acquisition

From **yahoo finance** (library - quantmod) We pulled daily closing prices (`X_price`) of futures contracts traded for Reformulated Blendstock for Oxygenate Blending, RBOB for short, a gasoline futures contract and the benchmark for wholesale gasoline (NYMEX: RB = F) and Heating Oil, a distillate fuel used for heating purposes (NYMEX: HO = F) going as far back as 2000.

1.2 Preprocessing

- **N/A Handling:** Fill missing values with the mean of the previous 7 days
- **Data Parity:** Transformed daily *closing prices* to logarithmic scale, to compare “apples to apples”
- **Daily returns :** Log returns calculated

Key Finding: Correlation of returns = **+70%**



Figure 1: Normalized Fuel Prices (Rebased at 100)

1.3 Time Series Modelling

1.3.1 Stationarity check:

Augmented Dickey–Fuller (ADF) to logarithmic prices and returns.

For both gasoline and heating oil log-prices, the ADF p-values is bigger than the conf interval (0.05), so we **cannot reject the null hypothesis** (ie, the series is a random walk, what we want in a financial series)

For log-returns, ADF p-values are below 0.05, indicating stationarity ($I(0)$). So, non-stational prices and statinal returns require then a cointegration testing.



Figure 2: Daily Log Returns (2018 onwards)

1.3.2 Cointegration

Pairst trading is all about **assuming our variables move together over time**. To check this, we perform an Engle-Granger, that is both easy and powerful for pairs of variables. If they indeed move together, the spread (residuals of the model) should move around a long-term average.

If gasoline's and heating oil's closing prices move together over time, we should find the following from the model:

- **(intercept)** → the base difference between the stocks
- **(slope)** : equilibrium hedge ratio: **How much does HO move when RBOB moves**
- **R² (fit quality)** – not used for cointegration
- **Residuals** → the spread (actual price - predicted price)

We find evidence of cointegration between RBOB and HO: the ADF test on the residuals returns $p = 0.001$, which we will consider significant for this analysis. We proceed with our trading implementation.”

1.3.3 AR(1) Model on the Spread

$$(\phi(\text{ar1}) = 0.9851)$$

This means:

The spread is mean-reverting, because $|\phi| < 1$

But it is not extremely fast (ϕ close to 1 → slower reversion). In comparison to equities that have a faster correction, commodities tend to revert slower.

1.3.4 Half-Life of Mean Reversion

An important metric is half-life, that calculated how long does the spread take to move halfway back to its long-run mean.

half-life = 46.1 days

If the spread widens today, it takes ~46 trading days. Commodities usually follow longer cycles of supply and demand and suffer less volatility than equities. Nevertheless, this is longer in comparison to equities, which typically have half-lives of 5–20 days.

1.4 Findings

Once we have calculated our spread and checked the existence of a relationship between gasoline and heating oil, we need to start setting windows to compare if in recent times the market has been bearish or too excited against its historical.

Starting with the rolling window, we must have enough history to trust our rolling mean and z value. The quantitative finance literature (Pole (2007), Gatev et al. (2006)) indicates that a good rule of thumb for the rolling window should be half life x 1.5. Therefore, with a half life of ~46 days, we set our rolling window for the previous **60 days**

For each day in the dataset, we calculate the mean spread, and the volatility (standard deviation). Then, our Z score is no more than the coeff of how many standard deviations away from the mean my current spread is.

1.4.1 Building the signals

In order for an algorithm to operate, it needs rules to know when and how to trade. This is the concept of signals: on our case, a signal is simply an instruction to buy, sell, or hold a position. We use the z-score to build our signals as follows:

- **+1** → open / keep a long spread position. Buy gasoline and sell HO when the spread is abnormally low
- **-1** → open / keep a short spread position. Sell gasoline and buy HO when the spread is abnormally high
- **0** → close position (stay flat) When the spread returns to normal, the trade must be closed.

Note: This model is trained to *only* hold one trade at a time. Therefore, if a signal is already active (long or short), the model will ignore any new signal until the position is closed (signal = 0).

1.4.2 P&L Calculation

Our model is worth the amount of money it can make. Therefore, we need to calculate the profit and loss (P&L) of our strategy. WE have three metrics to evaluate it:

- **Annualized return:** Average return per year **13.5%**
This is assuming the model is operating continuously with no interruptions, no delays, and reinvested profits. For a market-neutral (long-short) strategy that does not depend on market direction, **this is a solid result considering annualized result of major indices** roams around 10-12% during the same period
- **Annualized volatility:** Standard deviation of returns per year **25.0%**
This is the standard deviation of the returns.
- **Sharpe Ratio:** Risk-adjusted return (higher is better)

By far, the most important metric for evaluating a trading strategy. It is calculated as the ratio of annualized return to annualized volatility (return per unit of risk)

Sharpe Ratio = Annualized Return / Annualized Volatility 0.537

For a market-neutral (long-short) strategy that does not depend on market direction, this is a solid result in terms of returns. However, we require risk management improvements to be practical in a real-world setting.

1.4.3 Stop-Loss integration

As mentioned in the chapter before, we require risk management so that our strategy is applicable in real-world settings. To further improve our strategy, we implement a simple stop-loss mechanic to limit the downside risk.

A stop-loss triggers an order to automatically sell (or buy, depending on the position) a stock when its price reaches a predetermined price (or percentage loss), helping to limit potential losses by exiting the trade before further losses may occur.

In our implementation, the stop-loss is integrated directly into the signal generation loop using z-score thresholds of +3 and -3, which represent extreme deviations (approximately 99.7% confidence interval under a normal distribution, assuming the spread is roughly Gaussian). Specifically:

-For a **long spread position** (signal = 1, where we've bought RBOB and sold HO expecting the spread to increase): If the z-score drops below -3 (indicating the spread has narrowed adversely even further, against our bet), the stop-loss triggers an immediate exit by setting the signal to 0, closing the position to prevent deeper losses. -For a **short spread position** (signal = -1, where we've sold RBOB and bought HO expecting the spread to decrease): If the z-score rises above +3 (indicating the spread has widened adversely, against our bet), the stop-loss similarly forces an exit to 0.

This mechanism is checked daily alongside the normal exit conditions (e.g., when $|z| < 0.5$). By using z-scores tied to the rolling spread volatility, the stop-loss adapts to changing market conditions rather than a fixed price, making it more robust for commodities like these. In backtesting, it helps reduce maximum drawdowns, though it may occasionally cut short recovering trades if volatility spikes.

1.5 Conclusion

We have implemented a pairs trading strategy that leverages the mean-reversion properties of two closely related energy commodities: RBOB gasoline and heating oil futures, both traded on the NYMEX. By exploiting their strong historical correlation of approximately 70% in daily returns, driven by shared crude oil inputs, refining processes, and seasonal demand factors, we identified temporary divergences in their price spread through time series techniques, including stationary tests, Engle-Granger cointegration (with a significant ADF p-value of ~ 0.001 on residuals), and an AR(1) model revealing a mean-reversion half-life of 46.1 days.

Backtesting on data from the year 2000 gave decent results: 13.5% annualized return, 25% volatility, and a Sharpe ratio of 0.537. That's pretty good for a market-neutral setup, beating average stock market returns (like 10-12% for the S&P) without relying on overall trends. We added a basic stop-loss at ± 3 z-scores to cut losses, but it might close winning trades too early sometimes.

From this project, we have learned a lot about how to handle time series data and explore new trading strategies for our personal Investments. We also have now deepened our knowledge in the energy-markets and gained a broader understanding of the underlying dependencies.

All in all we are very satisfied with our project.