

PAIRS TRADING

What is Pairs Trading strategy?

Pairs trading is a market-neutral trading strategy that involves simultaneously buying one asset (or going long) and selling another related asset (or going short). The objective of pairs trading is to profit from the relative performance of these two assets while minimizing exposure to overall market movements. It's based on the idea that, over time, certain related assets will move in tandem, and when they diverge from their historical relationship, there may be opportunities for profit.

Here's how pairs trading typically works:

1. **Pair Selection:** Traders first select a pair of assets that are believed to have a strong historical correlation. For example, they might choose two stocks from the same industry or sector, or they could use pairs of related financial instruments such as two exchange-traded funds (ETFs) that track similar indices.
2. **Historical Correlation:** Traders analyze the historical price movements of the chosen pair to identify when the prices have diverged significantly. They look for instances where one of the assets has become overvalued relative to the other.
3. **Entry and Exit Points:** Once the divergence is identified, traders establish a long position (buy) in the underperforming asset and a short position (sell) in the outperforming asset. The idea is that the underperforming asset will catch up, and the overperforming asset will decline in value.
4. **Risk Management:** Pairs trading involves managing risk carefully. If the spread continues to widen instead of converging, traders may need to adjust their positions or cut their losses.

Pairs trading can be a market-neutral strategy, meaning that it aims to generate returns regardless of the overall direction of the broader market. It's often used by quantitative and algorithmic traders who rely on statistical models and historical data to identify suitable pairs and execute trades.

How to select a pair among given dataset?

We have used NIFTY 100 stock indices for this project, and the stock data to find the pair used is from June'17 to June'20, which is then back tested on period from July'20 to July'23.

After successful collecting stocks closing prices, we have performed the following tests to find a pair among them:

1. Pearson Correlation Test (Pretest):

We are focusing on identifying pairs of stocks that exhibit a strong correlation. Therefore, we have established a specific threshold value of 0.9. In practice, this means that we will only proceed with those stock pairs in our analysis where the correlation between them exceeds 0.9. This approach helps us conserve computational resources for subsequent testing and allows us to concentrate on the most highly correlated stock pairs for further investigation.

The Pearson correlation test, also known as Pearson's correlation coefficient, is a statistical method used to measure the strength and direction of a linear relationship between two continuous variables. It quantifies the degree to which two variables are related, with values ranging from -1 to 1.

The Pearson's correlation coefficient for a pair of stocks is calculated as follows:

$$r = Cov(X, Y) / \sqrt{Var(X) * Var(Y)}$$

where, r = Pearson's correlation coefficient

X = Daily closing price of first stock

Y = Daily closing price of second stock

$Cov(X, Y)$ = Covariance between daily closing price of stock pair

$Var(X)$ = Variance of daily closing price of first stock

$Var(Y)$ = Variance of daily closing price of second stock

After performing Pearson's correlation test, we will be left with pairs which are highly correlated, but to capture more insights about the relation between stocks, we will perform cointegration test.

2. Cointegration Test:

Cointegration is a statistical concept used to analyze the long-term relationship between two or more time series data sets. The primary purpose of a cointegration test is to determine whether there exists a stable, long-term relationship between two or more non-stationary time series.

Stock price data are often non-stationary, which means their statistical properties, like the mean and variance, can change over time. Non-stationary data can make it challenging to analyze relationships between variables. Cointegration suggests that two or more non-stationary time series are "integrated" in a way that the combination or linear combination of these series becomes stationary. In simpler terms, cointegrated time series move together in the long run, even if they might exhibit short-term fluctuations and divergences.

There are mainly two types two types of cointegration test:

- a) **Engle-Granger Test:** This test involves regressing one non-stationary time series on another. If the resulting residuals (the differences between the observed and predicted values) are stationary, it suggests cointegration between the two series. This test is relatively straightforward but is suitable for examining pairs of variables.
- b) **Johansen Test:** The Johansen cointegration test is a more advanced method that can handle multiple time series. It allows you to assess not only whether cointegration exists but also how many cointegrating relationships are present in a set of variables.

In cointegration test we consider two hypotheses:

Null hypotheses, H_0 : The given pair is not cointegrated.

Alternate hypotheses, H_1 : There exists cointegration between the pair.

Corresponding to the null hypotheses, a probability value (p-value) is calculated which suggests the chances of null hypotheses to be true. If p-value < 0.01, 0.05, 0.1, we can reject the null hypotheses with 99%, 95%, 90% confidence. In our test we have selected pair of stocks whose p-value is less than 0.01.

3. Hedge Ratio Test

After we have performed both correlation and cointegration tests, we will be left with few pairs of stocks, out of which we will choose one on the basis of hedge ratio.

In pairs trading, the hedge ratio, also known as the hedging ratio or beta, is a crucial parameter that determines the number of shares of each stock to hold in a pair of assets.

The hedge ratio is determined through a linear regression analysis of the cointegrated pair. Specifically, it is the coefficient of the cointegrated asset X in a linear regression model:

$$Y_t = \alpha + \beta * X_t + \varepsilon_t$$

where, X_t represents the value of the first asset at time t.

Y_t represents the value of the second asset at time t.

α is the regression intercept.

β is the hedge ratio.

ε_t is the residual, representing the difference between the actual value of Y and the value predicted by the regression.

B can be calculated as follows:

$$\beta = \frac{Cov(Y_t, X_t)}{Var(X_t)}$$

where, $Cov(Y_t, X_t)$ = Covariance of Y_t and X_t and $Var(X_t)$ = Variance of X_t

We choose the pair whose hedge ratio is closest to 1 because a pair having hedge ratio close to 1 has many positives such as:

- a) **Market-Neutrality:** A hedge ratio close to 1 means that the positions in both assets are roughly equal. This balance is critical for achieving market-neutrality because it ensures that the total exposure to the market is minimized.
- b) **Relative Price Movements:** If the ratio is close to 1, it means that the two assets are moving in sync, and any divergence from the historical relationship may present an opportunity for profit.
- c) **Profit Potential:** When the hedge ratio is close to 1, any deviation from the historical relationship between the assets is more likely to result in a profit because the positions are balanced. If the ratio is significantly different from 1, it can lead to unbalanced positions, making it more challenging to benefit from relative price movements.
- d) **Risk Management:** A hedge ratio close to 1 also simplifies risk management. It helps ensure that the overall portfolio is less affected by unexpected market movements. With balanced positions, you can better control your risk exposure.

Trading Strategy

The trading strategy we formed will be back tested on period from July'20 to July'23.

We implement our trading strategy in following steps:

1. Calculating Spread:

Spread calculates the difference between the stock prices of selected pair.

$$S_t = Y_t - \beta * X_t$$

Where, S_t = Spread

Y_t = Daily closing price of second stock

X_t = Daily closing price of first stock

B = Hedge Ratio from previous period

2. Calculating z-score:

After spread is calculated, z-score of spread is calculated with a rolling window with lookback period of 21 days.

z-score is defined as: $zscore = \frac{X - \mu}{\sigma}$

X represents the spread of past 21 days and μ , σ are mean and standard deviation of X respectively.

3. Entry and Exit Positions:

We could perform a profitable trade when z-score deviates from its mean value.

We will enter a position when z-score becomes greater than 1 or less than -1 and exit the position when z-score is again close to 0.

In a position, we long the underperforming stock and short the another, expecting that prices will come again to mean value as they are highly cointegrated.

The values of 1 and -1 are hyperparameters and these can be changed according to individual's preference of risk and returns.

4. Risk Management and Position Sizing

Risk Management:

- Set clear stop-loss and take-profit levels to limit losses and secure profits.
- Diversify your pairs to reduce risk. Don't put all your capital into a single pair.

Position Sizing:

- Optimize your position sizes. This can help maximize returns while controlling risk.
- Calculate position sizes based on your risk tolerance and the distance from the entry point to the stop-loss.

Results Obtained

The results of pair selection strategy and trading strategy, which are obtained are summarized below.

- **Selected Pair of Stocks:** ONGC and COALINDIA with a correlation coefficient of 0.925, cointegration p-value of 0.00538, and hedge ratio of 1.022.
- **Performance Metrics of trading strategy on selected pair of stocks**
 - Compound Annual Growth Rate (CAGR) = 18.595%
 - Cumulative Returns = 65.453%
 - Maximum Drawdown = -10.102%
 - Sharpe Ratio = 1.196
 - Frequency of trades = 66.048 per year