# Statistical Arbitrage
with
Pairs Trading

Yash Mailapalli- 200050160
Landa Jitendra - 200050068
Mullu Abhiram Naidu- 200070046
Ayush Raj Tarway- 200260012

Introduction to Stat Arb

• **Definition:** Statistical arbitrage involves exploiting pricing inefficiencies between related financial instruments using complex mathematical models.

• **Basic Principles:** It relies on the law of large numbers, aiming to achieve profit through numerous small, low-risk trades over time.

• **Importance in Financial Markets:** This strategy enhances market efficiency and liquidity by correcting price discrepancies.
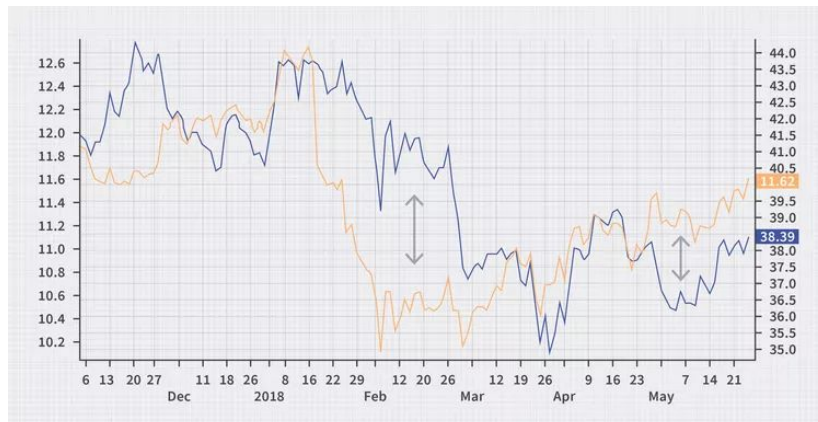
Pairs Trading and Cointegration

Concept of Pairs Trading: Pairs trading is a market-neutral trading strategy that matches a long position with a short position in two stocks with high correlation.

Selection of Pairs: Pairs are chosen based on historical correlation and statistical tests for cointegration to ensure they move together over time.

Goal of Pairs Trading: The strategy aims to capitalize on the relative price movements between the pair, profiting from convergence when they diverge.

Cointegration helps identify pairs of assets that move together in the long run, even if their short-term movements may diverge. This is based on the idea that certain assets or markets are linked by fundamental economic forces, such as supply and demand dynamics or industry trends.

Statistical Tests for Cointegration: Use of Johansen test and Engle-Granger two-step method to determine long-term equilibrium between pairs.

Formulation of Trading Signals: Trading signals are generated based on the mean reversion principle; if the spread widens, it is expected to revert back to its mean.

Identification of Cointegrated Pairs of Stocks

For this project we have taken the companies from NSE-100 which are categorized as 'FINANCIAL SERVICES' companies. After basic filtration we end up with 15 stocks. We split the closing price data for these stocks into test and train datasets.

We will use the Pearson correlation coefficient to get the basic idea about the relationship between these stocks and then work to identify cointegrated stocks using the function coint form statsmodels.tsa.stattools returning p-values of the cointegration test for each pair. We will store these p-values in an array and visualize it as a heatmap. The coint function uses the Engle-Granger (AEG) method to test for cointegration between two non-stationary time series variables.

The Engle-Granger (AEG) two-step method tests cointegration between two time series variables. First, it regresses one variable on the other to obtain residuals. Second, it tests these residuals for stationarity using unit root tests like ADF. Stationarity confirms cointegration, implying a long-term relationship between the variables.

Assets Correlation Matrix
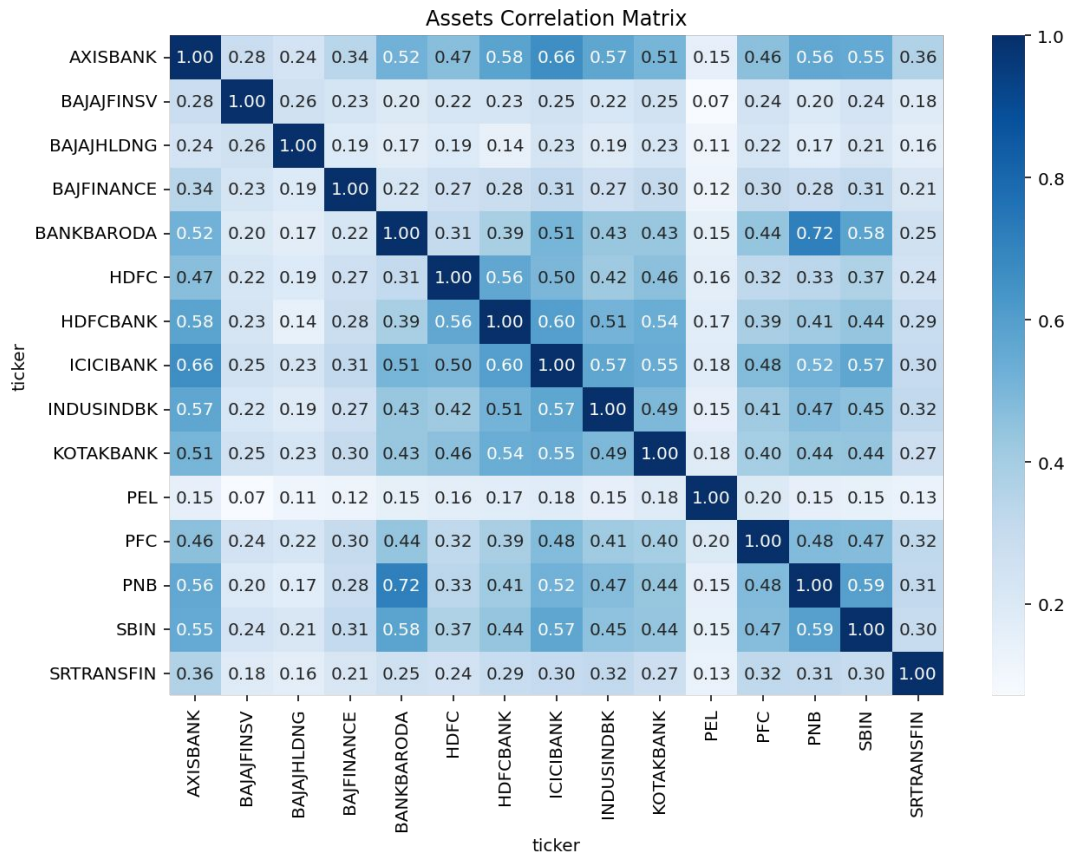
The value +1 indicates a strong positive correlation, zero indicates no relationship, and -1 indicates a strong negative relationship.

Assets Cointegration Matrix p-values Between Pairs
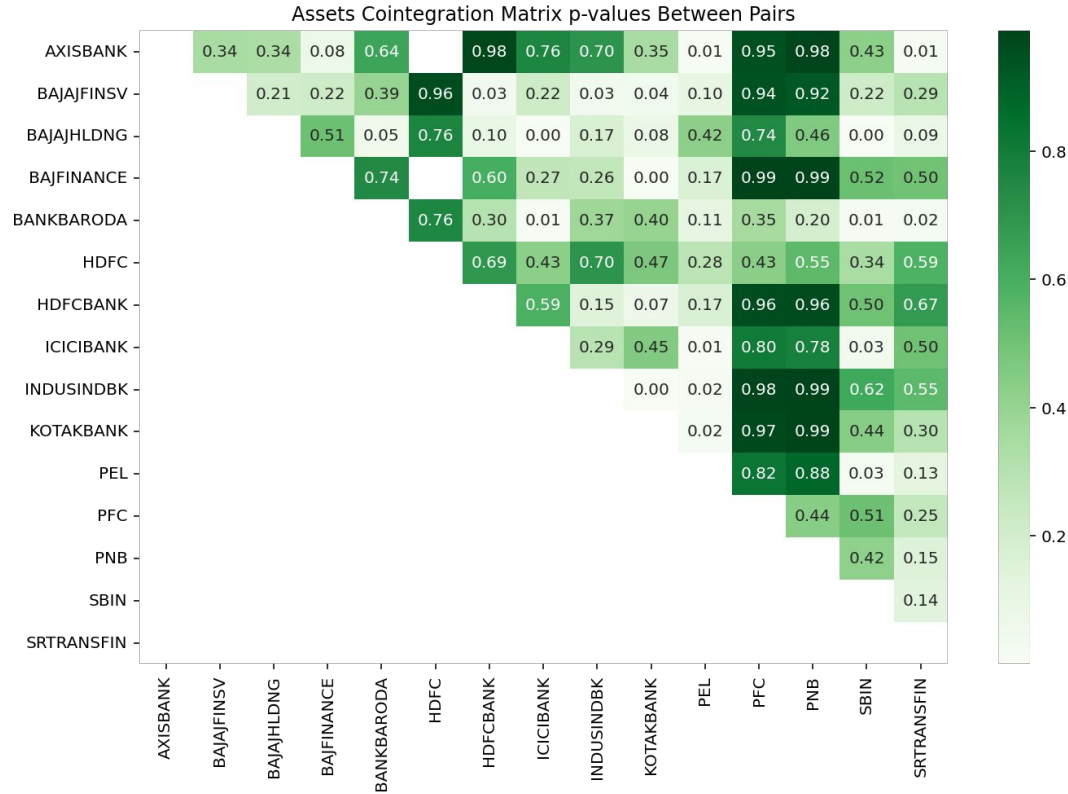
We can see in the above heatmap that there are many pairs with a p-value of less than 0.05. This means that for these pairs we can reject the null hypothesis and they can be cointegrated.

Perform Stationary Test for the Pair

Selecting the right pair is crucial for strategy success as the strategy will not work well if the prices are moving exactly together. (p-value less than 0.05)

Pairs need divergence and mean-reversion for profitability.

Stationarity of spread is essential.

We will first calculate the hedge ratio between these two tickers using OLS regression.

Then, using the hedge ratio, we will calculate the spread and run the Augmented Dickey-Fuller test.

Daily Closing Prices

Daily Closing Prices for BANKBARODA and SBIN

Daily Closing Prices for the Pair

Asset1 - BANKBARODA
Asset2 - SBIN

We can see from the above plot that closing prices between these two stocks move quite together.

OLS Regression

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                 asset2   R-squared (uncentered):              0.990
Model:                            OLS   Adj. R-squared (uncentered):         0.990
Method:                 Least Squares   F-statistic:                     1.333e+05
Date:                Wed, 17 Apr 2024   Prob (F-statistic):                   0.00
Time:                        18:08:15   Log-Likelihood:                    -5373.7
No. Observations:                1285   AIC:                             1.075e+04
Df Residuals:                    1284   BIC:                             1.075e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
asset1         1.4046      0.004    365.105      0.000       1.397       1.412
==============================================================================
Omnibus:                       23.562   Durbin-Watson:                   0.037
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               24.535
Skew:                           0.338   Prob(JB):                     4.70e-06
Kurtosis:                       3.011   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```
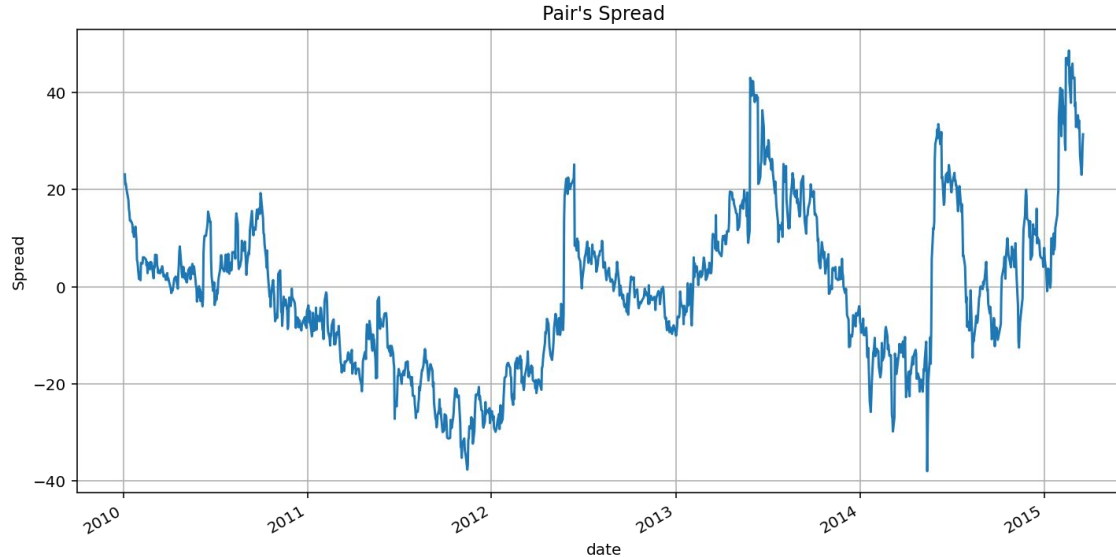
A high value of R-square and near-zero p-value from OLS regression suggest a very high correlation between these two stocks.

Pair's Spread

Pair's Spread

The spread looks stationary and the critical value from the Augmented Dickey-Fuller test is -3.459 which is less than the value at 1% (-3.435) significance level.

Hence, we are able to reject the null hypothesis that spread has a unit root and can conclude that it is stationarity in nature.

Generate Trading Signals Using Z-Score

We have used training dataset until now to finalise stock pair for our strategy.

But from now, we are using test dataset to ensure trading signal generation and out of sample dataset for backtesting

We use Z-score of the ratio between stock prices to generate trading signals and set upper and lower thresholds

By analysing the Z-score we are determining how far the price ratio is deviating from its historical average price ratio

A positive Z-score above the upper threshold indicates that the current price ratio is significantly higher than the historical average

- We interpret stock A price to decrease and stock B price to increase (price ratio = A/B)
- Therefore the strategy would generate the signal to short A and long B

Pair Trading – Trading Signals and Position

Pair Trading - Trading Signals and Position

We visualized both the stock prices along with its long and short positions in the portfolio.

Portfolio Profit and Loss Calculation

We will start with the initial capital of 100,000 and calculate the number of shares to buy for each stock.

On any given day, total profit and loss from the stock will be total holding in that stock and cash position for that stock.

We calculate holding in the stock by taking the cumulative sum of its position multiplied by stock price and the total number of shares.

We then calculate the cash position by subtracting holding from the initial cash position.

We follow the above steps for both the stocks and sum up two asset's positions for the total portfolio value.

Portfolio PNL

Portfolio Performance with Profit and Loss
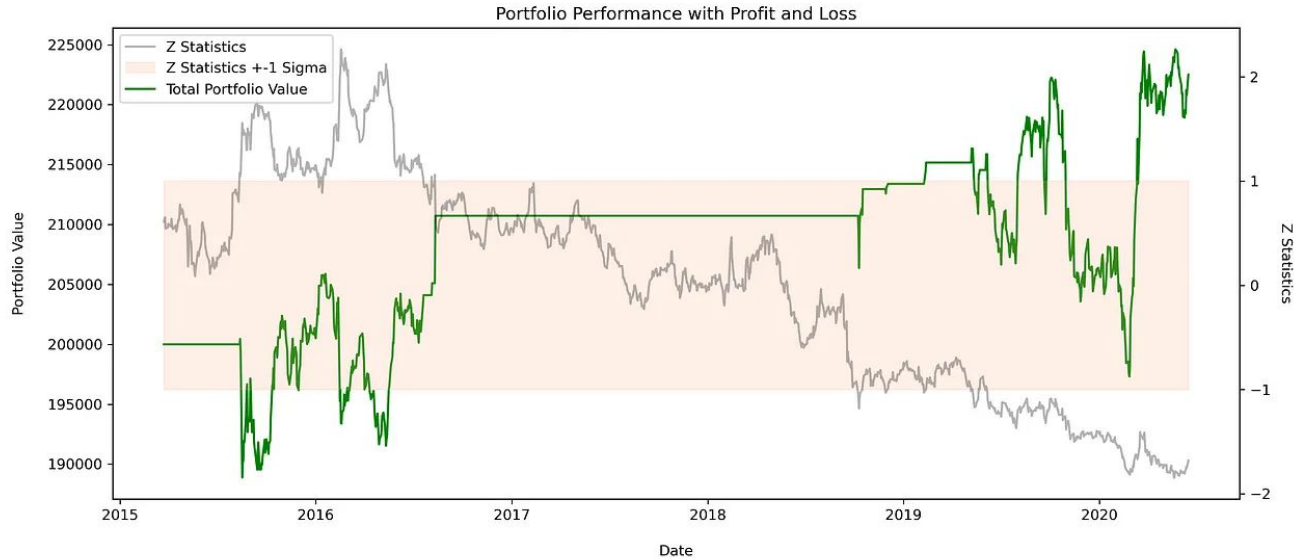
We have Visualized the portfolio performance along with z-score, upper, and lower thresholds.

Limitations and Conclusion

The Compounded Annual Growth Rate (CAGR) for the strategy is 16.5% which looks promising however there are many things to consider before we draw any conclusion. Few important factors to account for are as follows:

1. As this is a market-neutral approach a lot depends on our ability to short sell which may be limited due to various reasons including regulations

2. We have not accounted for costs related to trading, market slippage, and security borrowing. Normally, a market-neutral strategy results in a high number of trades.

3. There is always a limitation of using historical data to forecast the future.

Keep in mind that any decision to implement a strategy should be based only after considering all the critical performance parameters including its feasibility and returns net of fee and charges.