**Author-Jai Kaushik**

**ID-2024A3PS0419P**

**Date-11/6/2025**

**Topic-Statistical Arbitrage-Pair Trading Strategy**

**Pair Trading Strategy: A Quantitative Analysis and Implementation**

**1. Objective**

The primary objective of this project was to design, implement, and backtest a statistical arbitrage strategy known as "pair trading." This strategy aims to capitalize on the mean-reverting behavior of the price "spread" between two historically correlated assets(assets whose p value<.05 in the engle granger test ). The underlying principle is that if two stocks/ bonds/assets etc. typically move together, a temporary divergence in their prices may present a chance to profit , as the spread is expected to converge back to its historical mean.

**2. Methodology and Assumptions**

**2.1. Stock Pair Selection**

For this analysis, I selected two pairs of Indian equities based on their strong fundamental and sectoral connections, which suggest a tendency for price co-movement:

* **HDFC Bank (HDFCBANK.NS) & ICICI Bank (ICICIBANK.NS):** These institutions are prominent private sector banks in India, operating within a similar regulatory and economic environment.(p value=.025<.05)
* **TATA POWER (TATAPOWER.NS) & JSW ENERGY (JSW ENERGY.NS):** Both are significant entities in the Indian energy and electricity sector , influencing and being influenced by common industry factors.(p value=.0224<.05)

Daily adjusted closing price data for the period 2020-01-01 to 2024-12-31 (5 year period)was acquired using the yfinance library. Data preprocessing involved aligning the datasets by trading dates and handling any missing values by row-wise deletion to ensure data was clean and utilisable.

**2.2. Cointegration Testing**

An important prerequisite for a robust pair trading strategy is the presence of a stable, long term relationship between the asset prices, known as cointegration. I employed the Engle Granger two-step cointegration test (as mentioned in the task pdf) on the logarithmic prices of the selected pairs. This test involves:

1. Running an ordinary least squares (OLS) regression of one log price series against the other.
2. Performing an augmented dickey fuller (ADF) test on the residuals of this regression.

A p-value of less than 0.05 from the ADF test on the residuals was considered the criterion for establishing cointegration, indicating that the spread is stationary and mean-reverting. If cointegration was not detected, the strategy for that particular pair was not executed

**2.3. Spread and Z-score Calculation**

The "spread" between the two stocks was defined as the difference between their natural logarithms: Spread=ln(Stock1 Price)−ln(Stock2 Price) ,where ln is the log to the base e

To standardize this spread and quantify its deviation from its typical range, a Z-score was calculated. This involved computing a rolling mean and a rolling standard deviation of the spread over a specified window (e.g., 60 days,30 days, 90 days ): Z score=Rolling Std Dev(Spread)Spread−Rolling mean(Spread)​ /Rolling std dev(spread)

**2.4. Dynamic Trading Logic and Thresholds**

The core of the trading strategy leverages the Z-score for signal generation. A key enhancement in this implementation is the use of **dynamic rolling thresholds**. These thresholds are not fixed values but rather adapt to the recent volatility of the Z-score itself, calculated as a multiple of the rolling standard deviation of the Z-score over a separate window (e.g., 120 days).

* **Entry Conditions:**
  + **Short the Spread (Long Stock2, Short Stock1):** A position is triggered when the Z-score exceeds the dynamic\_entry\_threshold, this suggests that Stock1 is relatively overvalued compared to Stock2, or vice versa, expecting a likely mean reversion downwards.
  + **Long the Spread (Long Stock1, Short Stock2):** A position is triggered when the Z-score falls below the negative dynamic\_entry\_threshold,this shows Stock1 is relatively undervalued compared to Stock2, expecting an upward mean reversion.
* **Exit Conditions:**
  + **Mean Reversion Exit:** Trades are closed when the Z-score reverts towards its mean (zero) and crosses the dynamic\_exit\_threshold. This is the primary profit-taking mechanism.
  + **Stop-Loss Exit:** A critical risk management feature, a stop-loss mechanism is triggered if the Z-score moves adversely beyond a dynamic\_stop\_loss\_threshold. This limits potential losses if the spread continues to diverge against the trade direction.

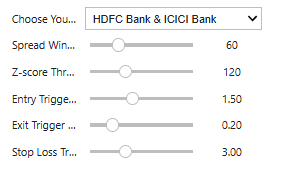
**2.5. Capital Allocation and Transaction Costs**

An initial capital base of 1,000,000 rupees was used for backtesting. For simplicity in capital allocation, an equal notional value was assumed for each leg of the pair trade. Transaction costs of 0.1% (0.001)(like mentioned in the task pdf) per side were incorporated for both entry and exit trades to simulate real-world trading expenses.

**3. Implementation of Bonus Tasks**

As part of the project, I implemented the following optional bonus tasks to enhance the strategy's functionality and analytical capabilities:

* **Rolling-Window Optimization for Dynamic Thresholds:** The core strategy was improved by replacing static entry/exit/stop-loss Z-score thresholds with dynamic ones. These thresholds are calculated as multiples of the Z-scores own rolling standard deviation (over a 120-day window) allowing the strategy to change according to the varying market volatility and spread behavior.
* **Application to Multiple Stock Pairs:** The entire pair trading strategy logic was packed within a reusable Python function (run\_pair\_trading\_strategy). This modular design facilitated the straightforward application and comparative analysis of the strategy across the stock pairs (e.g., HDFC Bank/ICICI Bank and TATA POWER/JSW ENERGY).
* **Interactive Dashboard (Jupyter Widgets):** To provide a dynamic and user-friendly interface for parameter exploration(tried streamlit but was not able to learn it) and result visualization, an interactive dashboard was developed using ipywidgets. This dashboard features:
  + **Dynamic Parameter Selection:** Sliders and dropdowns enable real-time adjustment of key strategy parameters, including the rolling window sizes for spread and Z-score threshold calculations, and the multipliers for entry, exit, and stop-loss Z-scores.
  + **Real-time Visualization Updates:** As parameters are modified via the widgets, the entire backtest re-executes, and all associated performance metrics and plots (Z-score trajectory, equity curve, comparative returns, trade P&L histogram) are updated instantly. This significantly aids in understanding the sensitivity of the strategy to different configurations.



**4. Performance Metrics**

The strategy's performance was rigorously evaluated using a comprehensive set of quantitative metrics:

* **Total Return:** The cumulative percentage gain or loss over the entire backtesting period. Cummulative return is just the return that gets added from the start date till the present
* **Annualized Return:** The geometric mean return expressed on an annual basis, providing a standardized measure for comparison.
* **Sharpe Ratio:** An important risk-adjusted return metric, calculated as the annualized return divided by the annualized standard deviation of daily returns (volatility), we can annualize volatility by multiplying with root of the number of turns . A higher Sharpe Ratio indicates better risk-adjusted performance.
* **Maximum Drawdown:** The largest peak-to-trough decline in portfolio value during the backtesting period, representing the greatest historical capital at risk.It is considered to be a good representation of risk , better than only volatility.
* **Number of Trades:** The total count of completed round-trip trades executed by the strategy
* **Hit Ratio (Win Rate):** The percentage of profitable trades out of the total number of completed trades
* **Average Holding Period:** The average duration that a trade remained open.

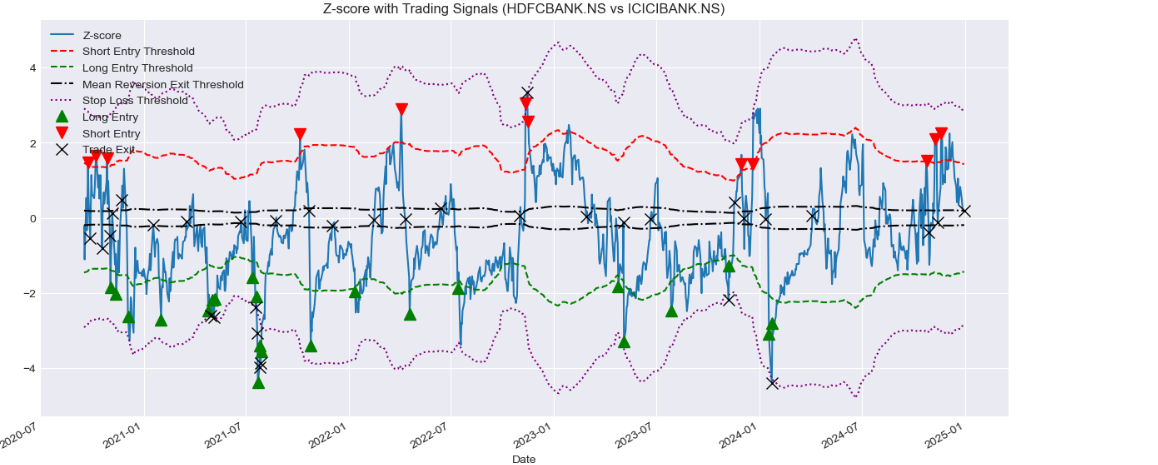
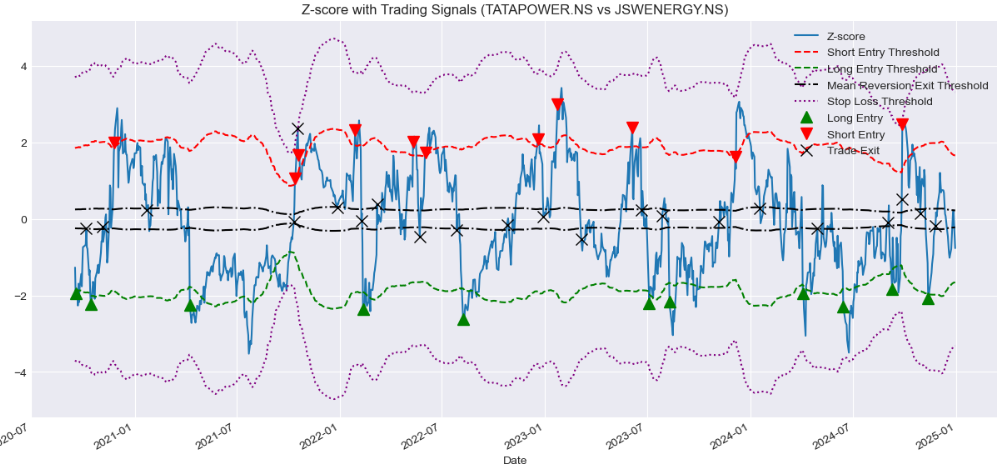
**Example Performance Metrics (for HDFCBANK.NS - ICICIBANK.NS with default parameters)**

*(Please note that these values are based on a specific set of default parameters and data period. The interactive dashboard allows for real-time calculation and visualization with user-defined parameters.)*

* **Total Return:** 23.03%
* **Annualized Return:** 4.96%
* **Sharpe Ratio:** 0.82
* **Maximum Drawdown:** -10.27%
* **Number of Trades:** 34
* **Hit Ratio (Win Rate):** 64.71%
* **Average Holding Period:** 26.53 days

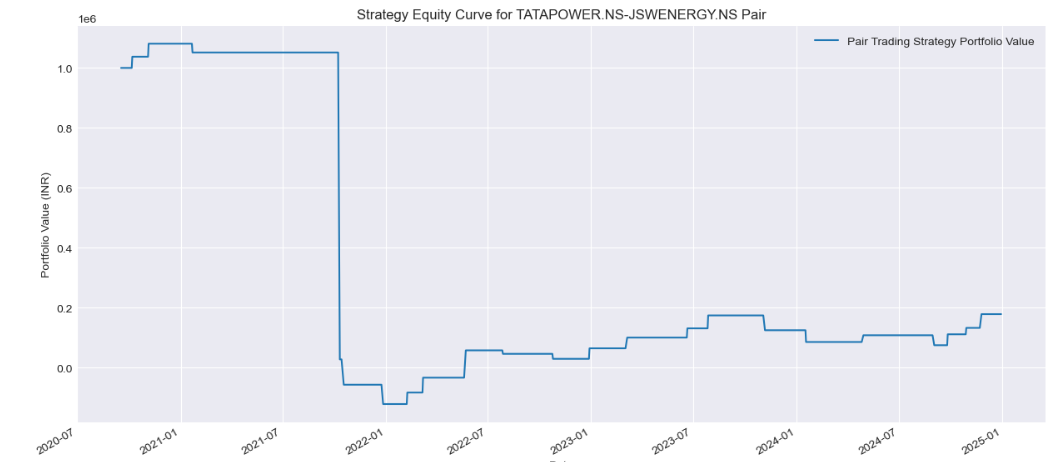
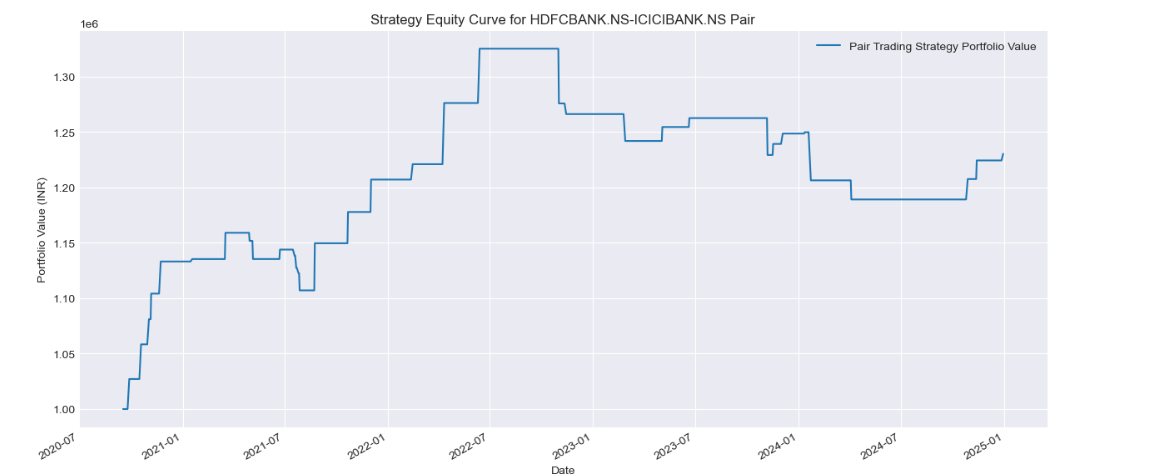
**5. Visualizations** The interactive dashboard provides the following visualizations to show the strategy's behavior and performance:

**5.1. Z-score with Trading Signals**

This plot displays the time series of the calculated Z-score for the spread. Overlaid are the dynamic entry, exit, and stop-loss thresholds, providing insight into the strategy's signaling mechanism. Trade entries (long and short) and exits are separately marked, allowing for isual correlation between Z score movements and trading actions 

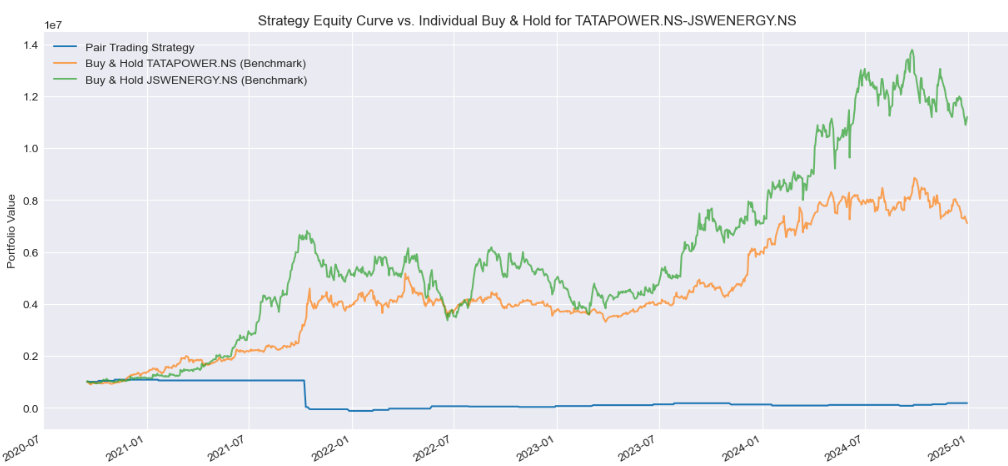
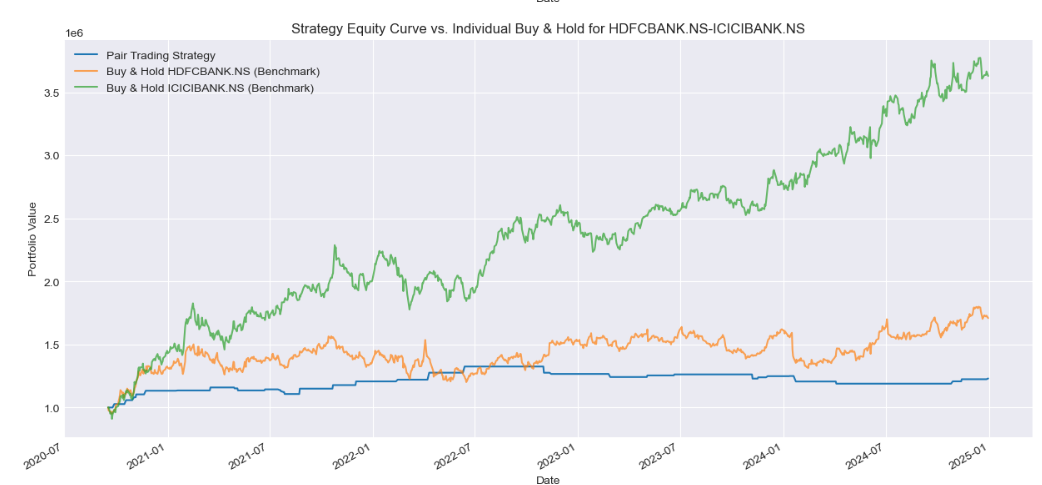
**5.2. Strategy Equity Curve**

This chart presents the cumulative portfolio value over the backtesting period. It serves as a fundamental visual representation of the strategy's growth trajectory and overall profitability.



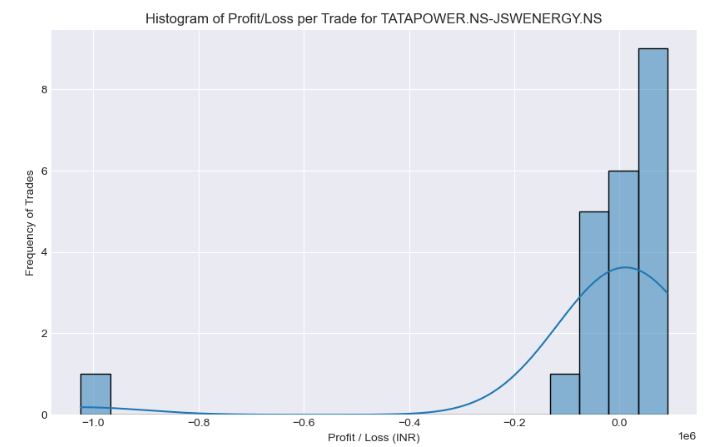
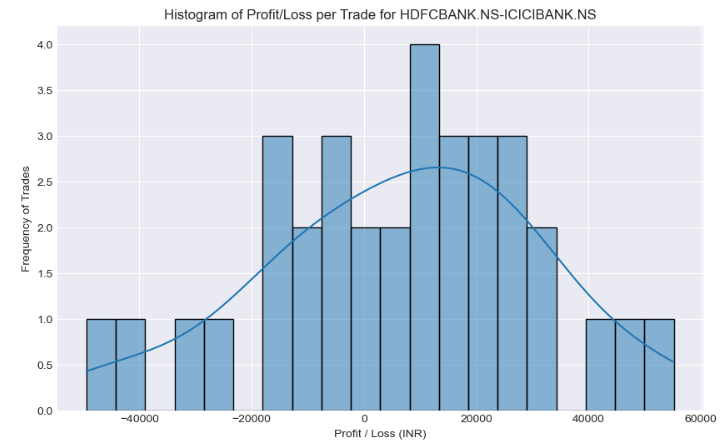
**5.3. Strategy Equity Curve vs. Buy & Hold**

This comparative plot illustrates the performance of the pair trading strategy against a simple buy-and-hold benchmark for each individual stock in the pair. This comparison is essential for assessing whether the strategy generates alpha (excess returns) relative to a passive investment approach.



**5.4. Histogram of P&L per Trade**

This histogram provides a distribution of the individual profit and loss outcomes for all completed trades. It helps in understanding the frequency and magnitude of profitable versus unprofitable trades, offering insights into the consistency and risk profile of the strategy's execution.



**6. Limitations and Future Enhancements**

**6.1. Current Strategy Limitations**

* **6.1. Current Strategy Limitations**
* **Reliance on Cointegration Stability:** A core assumption of this strategy is that the historical cointegration relationship between the selected pairs will persist. In practice, these relationships can and do break down due to evolving market structures or company-specific events
* **Fixed Rolling Window Periods:** While I have implemented dynamic thresholds, the specific look-back periods (rolling window lengths) for calculating the spread's Z-score and the Z-score's volatility are fixed. Optimal window lengths might vary in different market conditions or over longer time horizons
* **Simplified Stop-Loss Mechanism:** My stop-loss is based on a fixed Z-score deviation from the mean. A more sophisticated approach would likely involve incorporating dynamic volatility measures or focusing on a fixed percentage of capital at risk
* **Basic Transaction Cost Model:** The 0.1% transaction cost per side is a simplification. Real-world trading involves variable brokerage fees, exchange fees, and potential slippage, especially for larger trade volumes, which aren't fully captured here

**6.2. Potential Future Enhancements**

* Building upon this foundational work, I see several promising avenues for future research and development to enhance the strategy's robustness and performance:
* **Advanced Cointegration Analysis:** Exploring more sophisticated cointegration tests, such as the Johansen test, could provide a deeper understanding of multivariate relationships and potentially identify more stable pairs
* **Adaptive Parameter Optimization:** Developing algorithms (e.g., using machine learning or genetic algorithms) to dynamically optimize the rolling window lengths and threshold multipliers would allow the strategy to self-adjust to changing market dynamics.
* **Machine Learning for Signal Generation:** Instead of solely relying on Z-score thresholds, integrating machine learning models could offer more nuanced entry and exit signals by predicting future spread behavior based on a wider range of features
* **Volatility-Adjusted Risk Management:** Incorporating stop-loss and take-profit levels that dynamically adapt to the current market volatility (e.g., using Average True Range) could lead to more intelligent risk management.