# **QF603: Project Proposal (Group 1)**

**Group Members:**

Allen Lim

Ee Jing Michelle

Harshita Sachdev

Kwek Shi Chin

Lee Chengda

Phoo Pyae Hsu Myat

Zhao Geping

# Implementing Pairs Trading Strategy Between Cointegrated Assets

## 1. Introduction

Pairs trading is a market-neutral strategy that involves identifying two historically correlated assets, monitoring the spread between their prices, and executing trades when the spread diverges from its mean. This strategy capitalizes on the statistical arbitrage opportunities arising from temporary deviations in the price relationship between the two assets. In this project, we will analyze 150 stocks from the US, Hong Kong, and Indian markets, to identify cointegrated assets within their markets and develop a robust pairs trading strategy.

## 2. Objectives

The primary objective of this project is to develop and implement a profitable pairs trading strategy utilizing cointegrated assets. To achieve this, we will conduct rigorous statistical analyses to test for stationarity and cointegration. Additionally, we aim to identify optimal entry and exit points based on the dynamics of the spread between assets. The performance of the strategy will be evaluated through comprehensive backtesting, taking into account transaction costs and slippage to ensure realistic trading conditions.

## 3. Methodology

### 3.1 Data Collection

For this study, we will collect historical price data for the following assets:

* 4700+ stocks from US (NASDAQ, NYSE, AMEX)

https://www.nasdaq.com/market-activity/stocks/screener

The historical price data will be sourced from reliable providers such as Bloomberg, Yahoo Finance, or direct market exchanges. The time frame for analysis will encompass daily historical data over a ten-year period from 1 January 2014 to 31 December 2023 to capture long-term trends and deviations effectively.

Afterwards, the stocks will be grouped according to:

* Country of Risk
* Market Cap
* Sector

### 3.2 Cointegration Analysis

To identify stable long-term relationships among asset pairs, we will perform rigorous cointegration tests. The Engle-Granger test will be employed as a two-step method involving linear regression estimation followed by testing the stationarity of residuals using the Augmented Dickey-Fuller (ADF) test.

### 3.3 Spread Calculation

The spread () will be defined as the difference in the prices of cointegrated assets ( and ) with a hedge ratio (*β*).

Hedge ratio (*β*) is computed using OLS regression between the log price of the 2 assets.  
Alternative to OLS consider using:

* Ridge: more regularized and thus avoid the problem of overfitting
* Theil Sen: capable of handling corrupted data

We will analyze it for mean-reverting behavior which is essential for pairs trading. To test for mean reversion, we will use the ADF test. A low p-value (< 0.05) will indicate that the spread is stationary.

### 3.4 Trading Strategy

Our trading strategy will include specific entry and exit signals based on spread movements. We will enter a long position in one asset and a short position in the other when the spread deviates by a predetermined number of standard deviations from its mean. Positions will be closed when the spread reverts to its mean or reaches predefined profit or loss thresholds.

Parameters to optimize performance:

* Stop loss (how far from mean)
* Take profit (how close to mean)
* Entrance (number of SD from mean)
* Trade volume (per trade)

### 3.5 Backtesting

We will conduct extensive backtesting of our strategy using historical price data over the past decade. The performance evaluation will include key metrics such as Sharpe ratio, maximum drawdown, and cumulative returns to assess profitability and risk exposure accurately. Transaction costs and slippage will be incorporated into simulations to reflect realistic trading conditions.

## 4. Statistical Tests

To detect cointegration and assess the viability of our pairs trading strategy, we will employ several statistical tests:

* **Ridge Test:** OLS model seeks to find the coefficients that minimize the mean squared error. On the other hand, Ridge Regression tries to find the coefficients that minimize the mean squared error and wants the magnitude of coefficients to be as small as possible. That means each feature should have a little effect on the outcome. Therefore, Ridge Regression will perform worse than the ordinary least squares model on the training set. But it will give us better regularization and performance on the test set.
* **Thiel-Sen Test:** Compared to the OLS (ordinary least squares) estimator, the Theil-Sen estimator is robust against outliers. It has a breakdown point of about 29.3% in case of a simple linear regression which means that it can tolerate arbitrary corrupted data (outliers) of up to 29.3% in the two-dimensional case.
* **Engle-Granger Test:** This two-step method involves first estimating a linear regression between the two asset prices and then testing the residuals of this regression for stationarity using the Augmented Dickey-Fuller (ADF) test. If the residuals are stationary, the assets are considered cointegrated.
* **Augmented Dickey-Fuller (ADF) Test:** This test checks for stationarity in the residuals after regression analysis. A low p-value (typically < 0.05) indicates that the null hypothesis of non-stationarity can be rejected, suggesting cointegration.

In terms of hypothesis testing, our null hypotheses state that there is no cointegration between selected asset pairs and that the spread does not exhibit mean reversion. Conversely, our alternative hypotheses state that there is evidence of cointegration and that the spread is indeed mean-reverting. Statistical significance will be tested through 10%, 5% and 1% confidence intervals, guiding our inferences accordingly.

## 5. Performance Evaluation

### 5.1 Key Performance Indicators (KPIs)

We will measure success through several KPIs:

* Profit and Loss (P&L): Total returns generated by the trading strategy.
* Sharpe Ratio: Measure the risk-adjusted returns.
* Maximum Drawdown: Assess the largest peak-to-trough decline.
* Win-Loss Ratio: Reflects trade success rates.

### 5.2 Sensitivity Analysis

To ensure robustness, we will conduct sensitivity analyses by varying key parameters such as entry thresholds (±1, ±2, and ±3 standard deviations), holding periods (5, 10, and 15 trading days), and transaction costs (0.1%, 0.2%, and 0.5% per trade), exit trades and stop losses. This will help us understand the impact of these parameter changes on overall strategy performance and identify the most optimal settings.

## 6. Conclusion

This proposal outlines a comprehensive plan for developing and implementing a pairs trading strategy utilizing assets from diverse markets including those in the US, Hong Kong, and India, alongside currency pairs. By leveraging robust statistical tests, we aim to identify meaningful cointegrated pairs while capitalizing on arbitrage opportunities through an effectively structured trading strategy. Through thorough backtesting and performance evaluation, we expect to validate both the profitability and robustness of this strategy across different market environments.