



# PAIRS TRADING & STATISTICAL ARBITRAGE – MEDIUM FREQUENCY TRADING

Statistical arbitrage is a profitable strategy. It is more profitable in markets with less competition. Therefore, the strategy is more effective on markets with less sophisticated investors.

In this exercise, we would like to explore the possibility of utilizing statistical arbitrage strategy in the China stock market.

## Task 1

We would like to store the historical daily data of equities trading in the Shanghai Stock Exchange (SSE) into a database and conduct analysis of it.

## Objectives

Develop a system for fetching and storing the complete daily historical data of all stocks traded in Shanghai stock exchange (SSE) since 2010 with the following technical requirements.

- 1) Use MongoDB to store historical data.
- 2) Create the following REST API to expose MongoDB data to users
  - a. Given start date, end date, ticker and desired fields (e.g. high, low, volume), return the corresponding time series of the given day range with the requested fields.
- 3) Create a script to keep updating the data in the database.

## Suggestions

- 1) Yahoo Finance has a comprehensive dataset of daily data.
- 2) Please beware that the database and the REST API development will be extended in future tasks. So, pay attention to how you manage and develop them.

## Task 2

After completing the development of the data system, it is time to conduct analysis of the data to explore possibilities of statistical arbitrage on the China market.

The subtask below serves as a guide on devising a statistical arbitrage strategy. You may use your own approach if the method is sound.

There are plenty of materials online on how to devise a statistical arbitrage strategy (samples attached toward the end). Those online materials could serve as the starting point on your idea.

### Subtask:

**A**

**Build a correlation metrics/model between stocks.**

**B**

**Build a clustering model to group stocks according to their correlations.**

**C**

**Given the cluster and correlations, devise a mean-reversion strategy.**

**D**

**Backtest your strategy.**

**Please document the method with appropriate formulas explanations for your step A, B & C. Also, create a slide to briefly summarize subtask A to C, and detailed slides on subtask D.**