**Statistical Arbitrage of Equity Futures on the NSE**

**Abstract**

This project attempts to build and back-test a very simple statistical arbitrage strategy on equity futures listed on the NSE. The equity futures are selected amongst the biggest firms in India by market capitalization from six different sectors. Thirty-two names across various sectors are selected to maximize any chances of catching a mean reverting spread between them.

**Data Gathering**

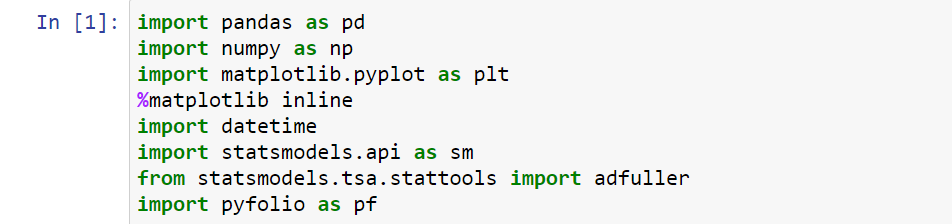
The data was collected through a subscription from an online broker. Data from 1st January 2018 till 22nd October 2020 was downloaded into a csv file into a local machine. The daily continuous contracts prices were downloaded from the broker website. These prices are considered for strategy development and back-testing. The csv file containing the data is attached along with the project report.

**Project Motivation**

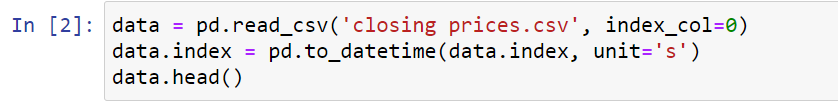
This project will serve as an excellent starting point to trade pairs trading strategy on the Indian Equity Futures market. At present, holding a short position overnight for many days is not allowed in the Indian markets. Therefore, this strategy is only suitable for futures contracts on stocks for which the positions can be rolled over into new months upon contract expiration. Mean reversion strategies such as this one are simple and robust to implement for retail traders because of their superior returns with respect to the risk even in highly volatile markets such as the first quarter of 2020 as we will see below.

**Strategy Development**

A few necessary libraries are imported into the code for data processing, time indexing, plotting and for the ADF test (Augmented Dickey Fuller Test). The PyFolio package is used to calculate and plot all the metrics



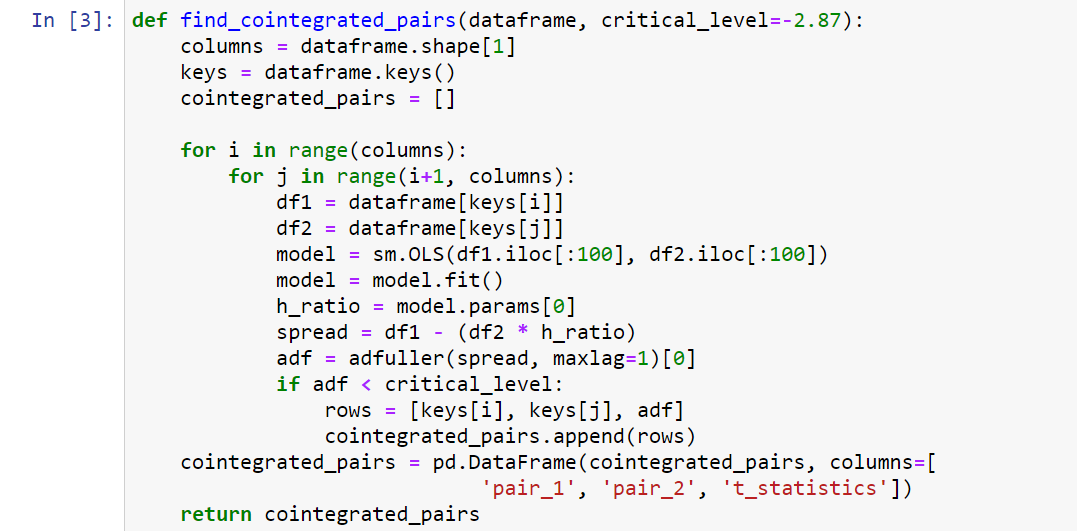
The libraries are imported. After that the main dataset is loaded into the program from the csv file. The dataset contains the time format in the UNIX timestamp format, so the necessary modifications have to be done to the dataset to convert to date-time format and setting the index to the date by applying the necessary Pandas functions.



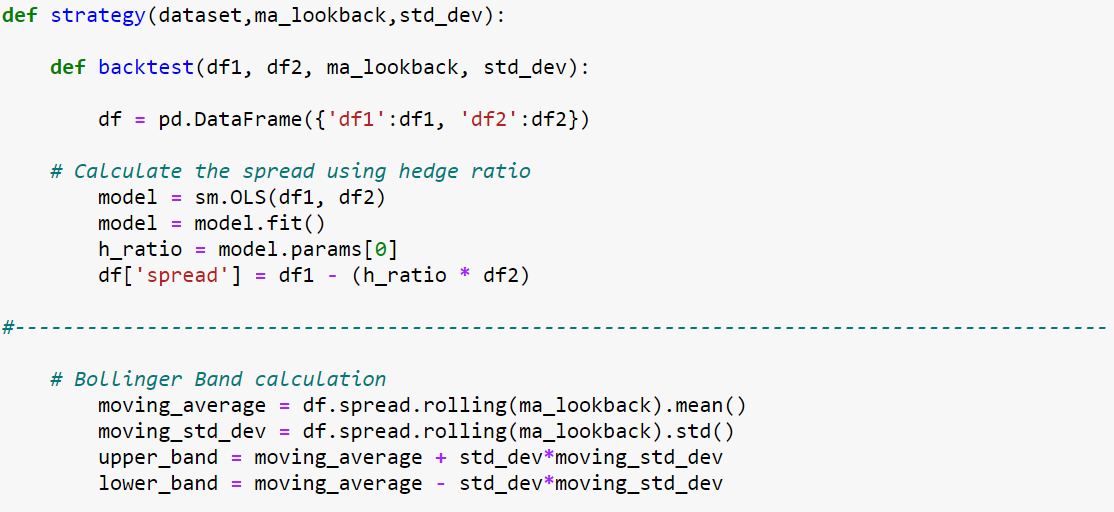
Here we have the data-frame head consisting of daily closing prices of 32 equity futures starting from 1st January 2018:

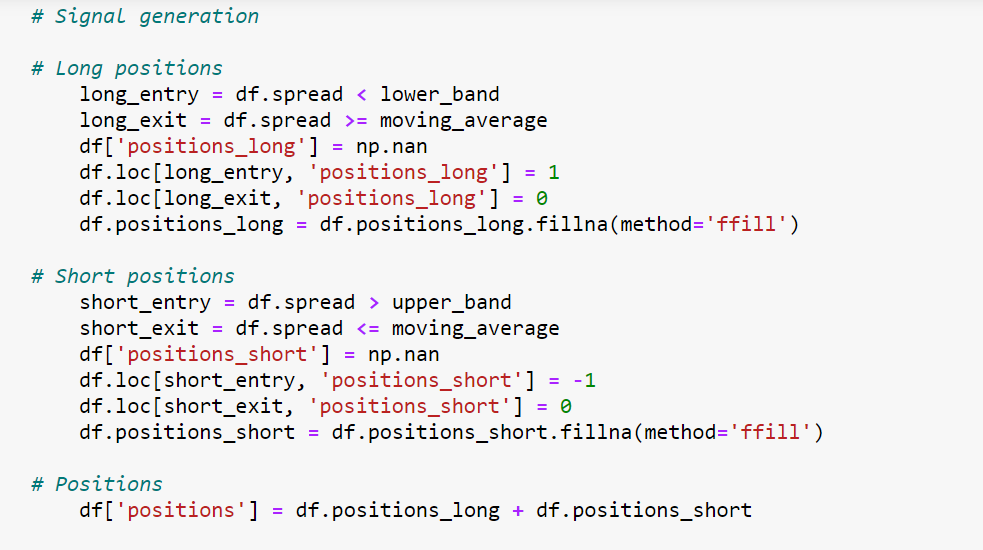


To provide modularity to the code, a function is defined as find\_cointegrated\_pairs with the data-frame and critical level as arguments. The spread is simply calculated amongst all the equities and the stationarity is checked and appended in the cointegrated pairs list for the critical value which in this case is set to -2.87 or 95% confidence that the spread is cointegrated:

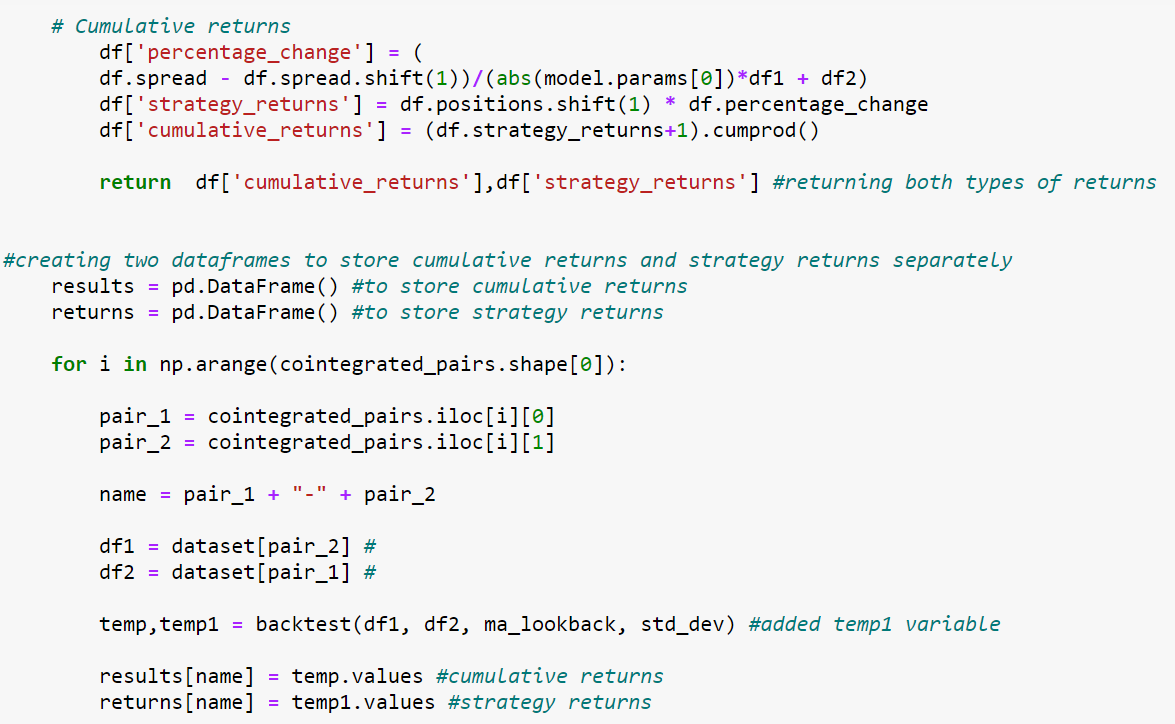


Finally, the spread between the selected cointegrated pairs is passed to the strategy function with arguments as the dataset, the moving average lookback period and the standard deviation which are required for calculation of the Bollinger bands. These arguments are used to run the strategy on in sample and out of sample testing with the various parameters for in-sample dataset parameter optimization:

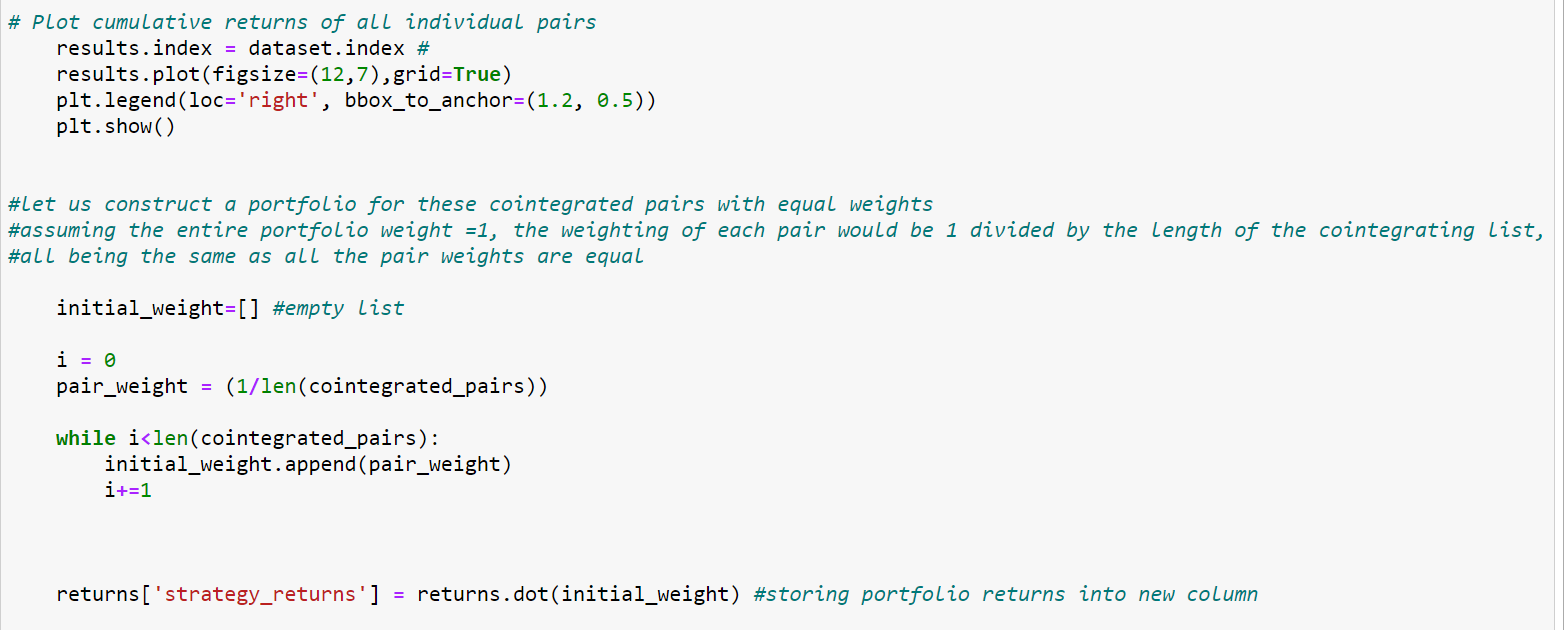




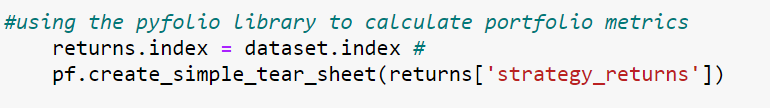
The backtest function calculates the buy and sell signals and calculates the cumulative and strategy returns which are used for plotting the returns of individual pairs in the backtest. The plots will be shown later.



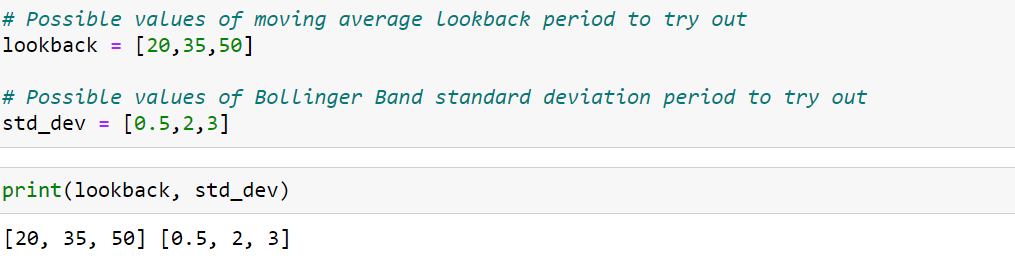
To calculate the portfolio returns, each of the individual pairs are equally weighted and multiplied by their corresponding strategy returns by the returns.dot() function which uses matrix multiplication and appends the strategy returns column with the daily portfolio returns.



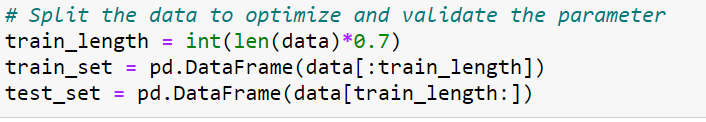
Finally, the strategy returns column is passed to the pyfolio library function of create\_simple\_tear\_sheet which outputs all the metrics of the portfolio.



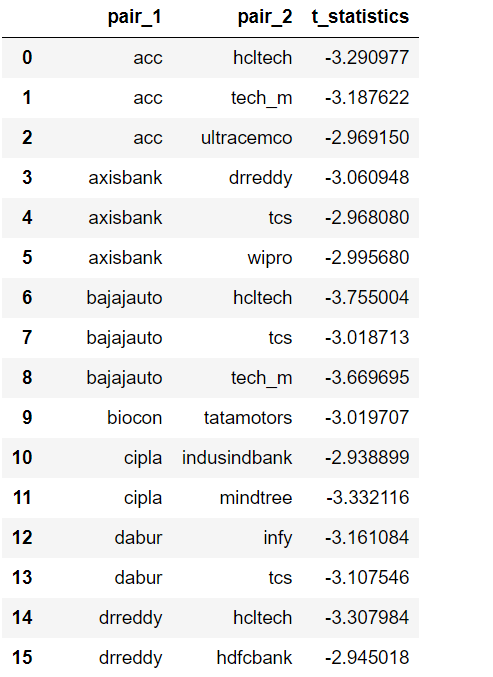
Now, we will try different values and combinations of moving average lookback and standard deviation for Bollinger Band to check which combination produces the best performance in the in sample data.

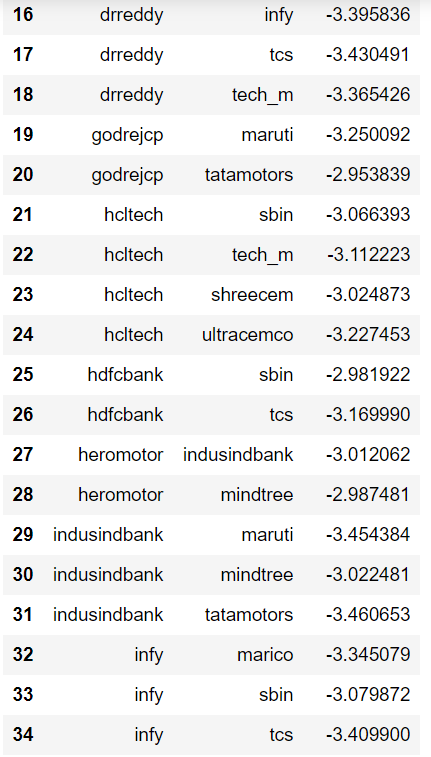


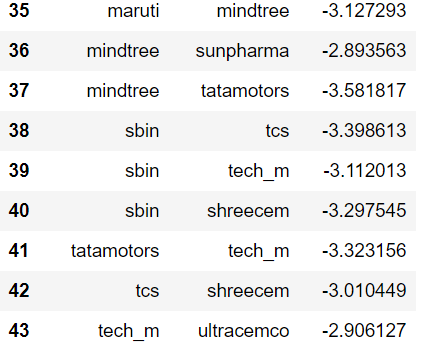
The dataset is split into the training set and the test set as per the code as shown. The training dataset is 70% of the total data and the test set is the remaining 30% of the dataset.



Now, the train dataset is passed to the find\_cointegrated\_pairs function to calculate all the cointegrated pairs in the training dataset with the 95% confidence which corresponds to critical level of -2.87. The pairs are stored in a data-frame with their T-Stat values.

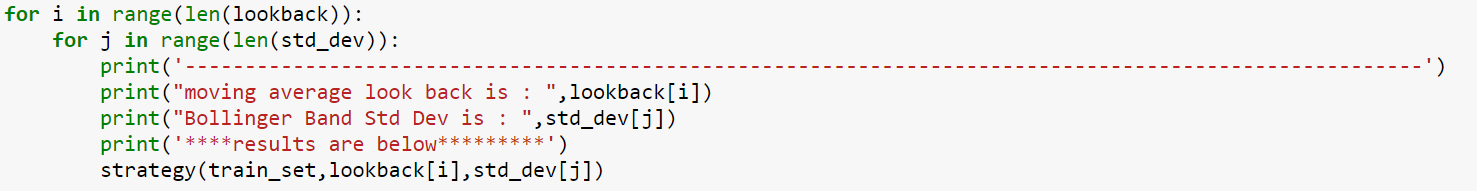






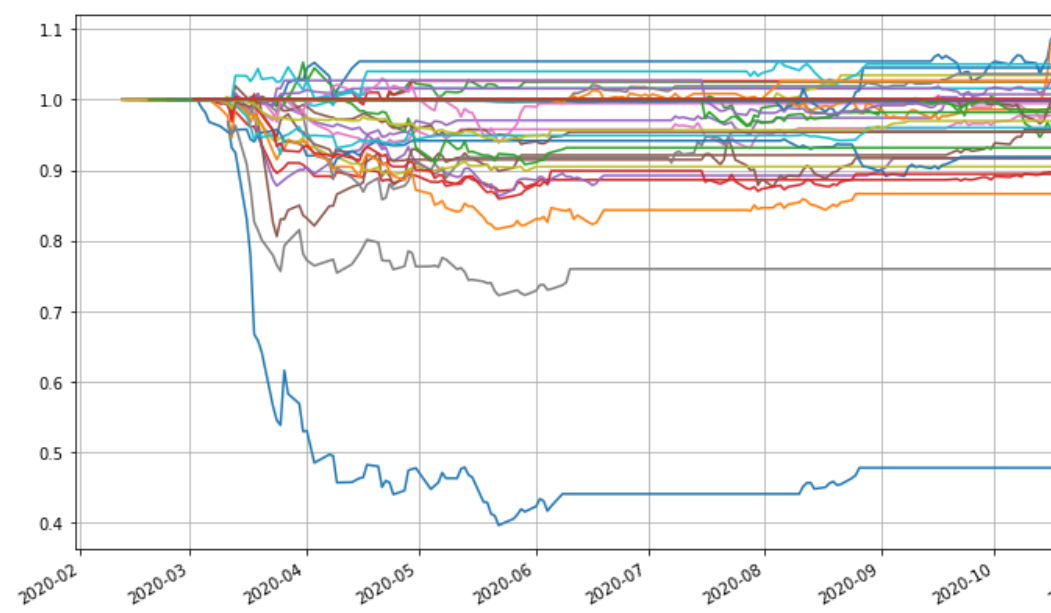
The data-frame contains 43 pairs.

Next, a For-Loop is used to try out all the combinations of MA-lookback and Standard deviation as per the values defined in the list above. The loop calls the main function strategy() by iterating over all the values stored in the lookback and std\_dev lists.

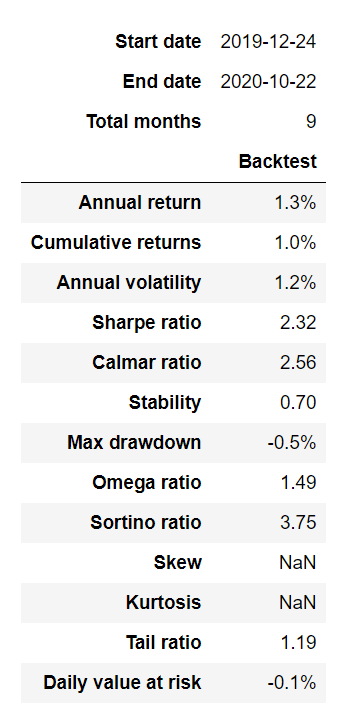


The code outputs the results for all the combinations. However, due to space constraints the outputs will not be shown here. The best performing portfolios of the parameters were found to be of Moving Average lookback of 35, Standard deviation of 3 and the portfolio for MA lookback 50 and standard deviation of 2 on the training data set.

Upon running the above parameters on the test dataset, the best performing portfolio was found to be of MA lookback 35 and Standard Deviation of 3. The performance of the individual pairs is shown below:



As seen above, the returns of the individual pairs suffer brutally during the market crash of the first quarter of 2020. However, when we calculate the portfolio metrics, the performance is as below:



In the portfolio, the annual and cumulative returns are low but the Sharpe ratio is pretty good at 2.32. Therefore, combining all the pairs into a portfolio and trading this strategy as a whole offers solid returns with respect to the risk.

**Conclusion**

1. Mean reversion strategies can suffer terrible drawdowns during market crashes as during the first quarter of 2020. However, combining lots of pairs into a portfolio greatly minimises the drawdown and offers excellent risk adjusted returns.
2. The above strategy did not consider the transaction costs and rollover costs associated with futures contracts, therefore these need to be taken into consideration when taking the strategy live.
3. As the strategy was traded on the Equity Futures market, leverage can be easily applied to magnify the annual returns as the Sharpe ratio is quite good.