

PAIR TRADING STRATEGY MODEL

BACKGROUND:

In this project, you are provided **with daily close price** of the following two stocks and their **corresponding FX (2016 - 2019)**.

Ticker	Name	Currency	Country	Exchange
RIO AU Equity	RIO TINTO LTD	AUD	Australia	ASX
RIO LN Equity	RIO TINTO PLC	GBp*	Britain	London

*1 GBp = 1/100 GBP

Goal is to come up with a trading strategy from the datasets.

APPROACH

1. To handle the data:

- Firstly, converted 'GBp' to 'GBP' in the London data
- Removed the dates that were not there in either one of the datasets or not there in the Forex conversions
- Converted both the datasets to prices in USD for that particular day for better development of the model.

Model:

- Split the datasets into 60-40 train-test ratio in order to make sure the model works for fresh data
- The rationale behind the model is to use **Pair Trading** which is a **Market Neutral Statistical Arbitrage Strategy**. We see that the two stocks are of the same company, so pair trading would be useful.
- In order to get this working, we first check the **correlation** of the prices of the datasets (> 0.9).
- Now, the important part is that time series data of stocks is generally correlated because of a trend, so we need to check if there is **co-integration between the stock prices**. That means, even though the stock prices are non-stationary, a linear combination of it is stationary.
- Given that the linear combination of the stock prices is stationary, we run a regression to find the predicted London Stock Price from the Australian one.
OLS formula: $\text{london_price} = \text{beta} * \text{aus_price} + \text{residual}$
- Then we trade on the residual of the OLS regression, because we know that the **residual will oscillate around the mean since the stocks are co-integrated**.
- Each time we trade, **we create a perfectly hedged trade**, with 100 London Stocks and taking the other position in the equivalent amount of Australian Stock for the day.

- h) The model assumes **no transaction cost** for trades and being able to **buy fraction number of stocks** to perfectly hedge the position

2. To validate the model, we split the data into training and test sets.

For each set, first we check:

- a) If the **correlation** between the London and Australian stocks is high enough (> 0.9)
b) We check with a significance level of 10%. Null hypothesis is that there is no co-integration.

We use the Engle and Granger approach built in in python to check for co-integration. We only move ahead by rejecting the null if the p value < 0.1 , since co-integration is an important factor.

- c) We check if the **OLS regression** run on the stock prices is statistically significant
 $|t| > 1.65$, at the 10% significance level.

Y -> London Stock

X -> Australian Stock

We use **robust** Standard Error to account for heteroscedasticity

OLS formula: $\text{london_price} = \text{beta} * \text{aus_price} + \text{residual}$

- d) We check if the **residuals are normally distributed**. Null Hypothesis being that the residuals are normal.

We only move ahead by failing to reject the null if the p value > 0.1 , since only then can we use the Z-score

```
Training set
Co integration check: True
P value: 0.000108627093269

Correlation Coeff: 0.994027035021
OLS t-statistic: 864.586593585

P-value for normality check: 0.468346327566
Residuals are normally distributed: True
```

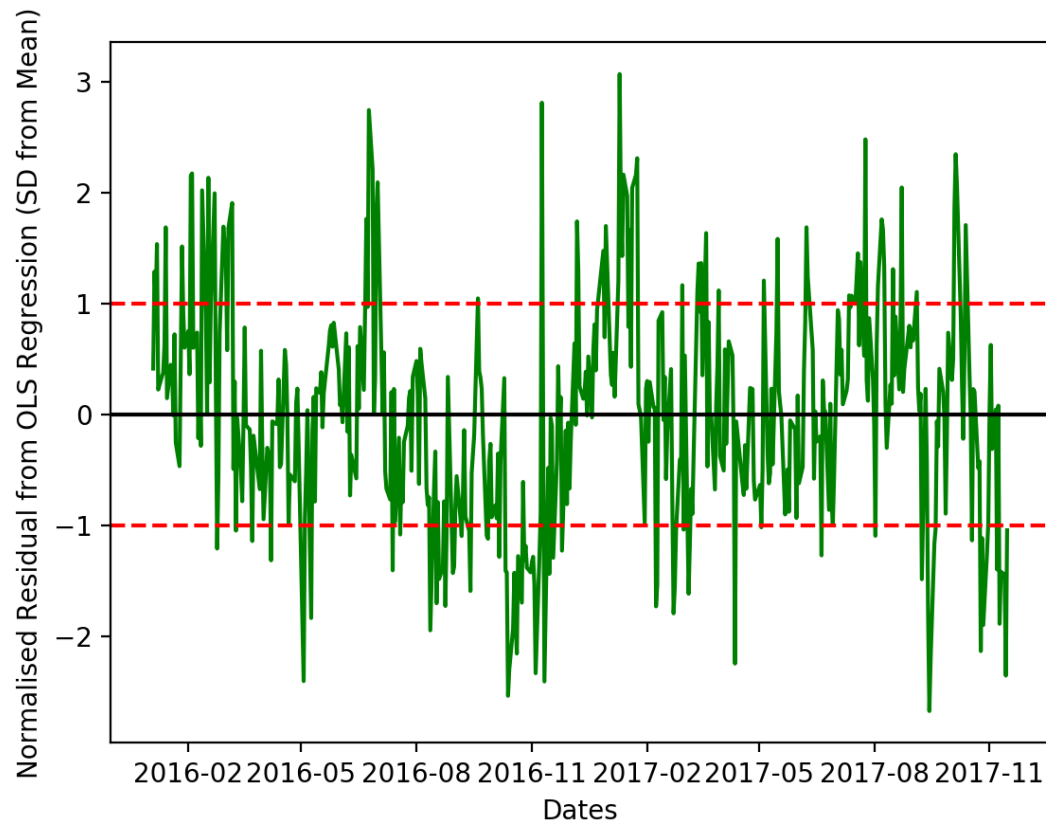
```
Test set
Co integration check: True
P value: 0.0988375144002

Correlation Coeff: 0.944560345704
OLS t-statistic: 618.300602053

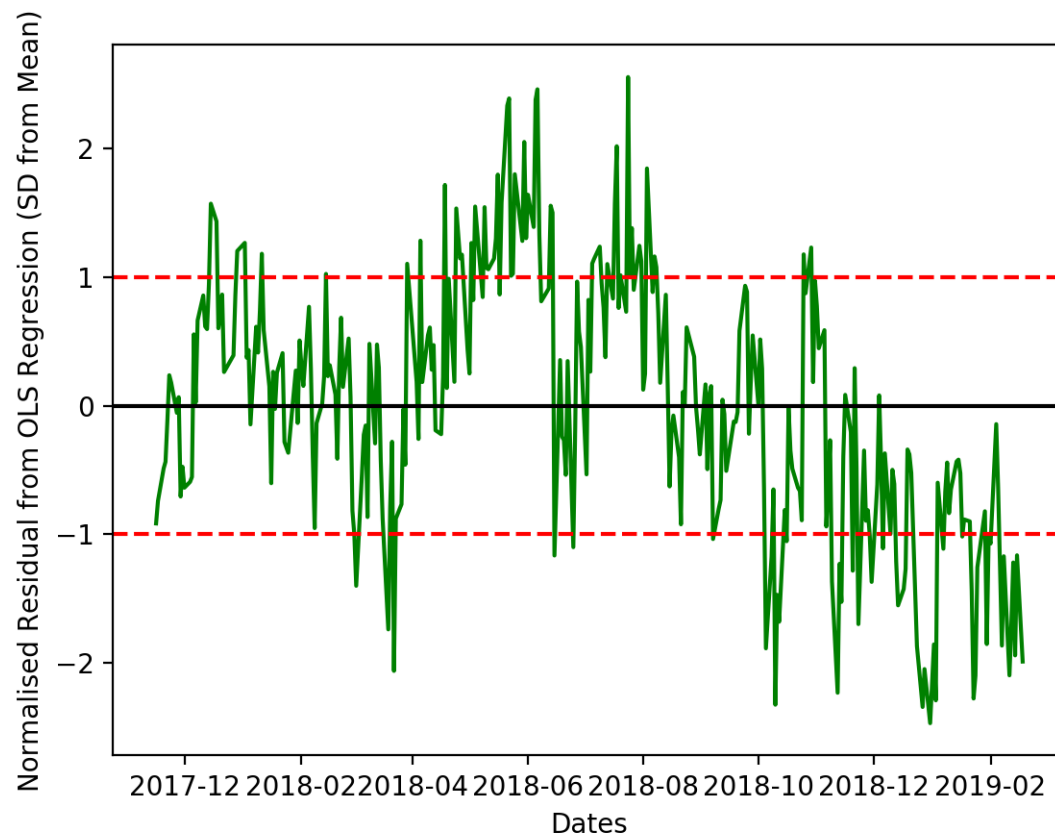
P-value for normality check: 0.530123526821
Residuals are normally distributed: True
```

- e) Furthermore, we graph the **residuals oscillating around the mean** to come up with the parameters used to trade (mentioned below)

Training Set



Test Set



3. The parameters in my strategy are:

a) **To Long London stock and Short Australian:**

If the **residual is less than 1 standard deviation below the mean**, that means that **London stock is undervalued or Australian is overvalued**.

Importantly, we create a perfectly hedged trade, where we long 100 London Stocks, and short equivalent ratio of Australian stocks

b) **To Short London stock and Long Australian:**

If the **residual is more than 1 standard deviation above the mean**, that means that **London stock is overvalued or Australian is undervalued**.

Importantly, we create a perfectly hedged trade, where we short 100 London Stocks, and long equivalent ratio of Australian stocks

c) **Exit the position:**

When the residual is **between 0.25 SD of the mean**.

d) **Stop Loss:**

To prevent humungous losses and not take on extra risk, we exit our position if it is **more than 3 SD on either side of the mean**.

e) **To minimize risk:**

Each time we trade, we create a perfectly hedged trade, with 100 London Stocks and taking the other position in the equivalent amount of Australian Stock for the day.

Apart from stop loss, I am ensuring that we do not take a position of more than 1000 London stocks at any given point of time in order to minimize the risk to our portfolio.

To come up with these numbers, I mainly used a graph to see how much the residual is oscillating around the mean, along with some trial and error on the training set

4. Simulated Results

In the pair trading strategy I am using, for every long and short, **the amount is perfectly hedged**.

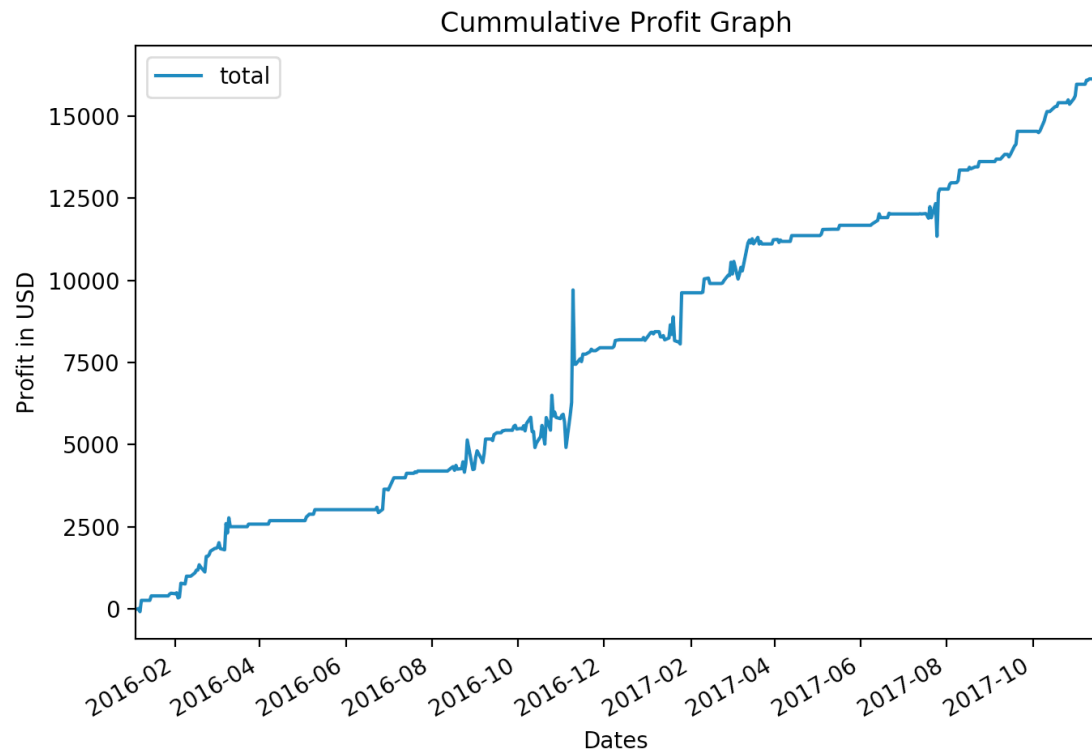
In order to calculate the capital required, I am assuming **50% margin** in the margin account.

So, the amount of capital would be the amount needed to long the selected stock and 50% of the amount we have shorted; which would technically be **1.5 times the amount needed to long the stock since our positions are perfectly hedged**.

Thus, the **investment required would be the maximum capital needed in the accounts at any point of time**.

Below attached is the profit graph that is being generated for the strategy where according to the signals mentioned in point 3 above.

Training Set

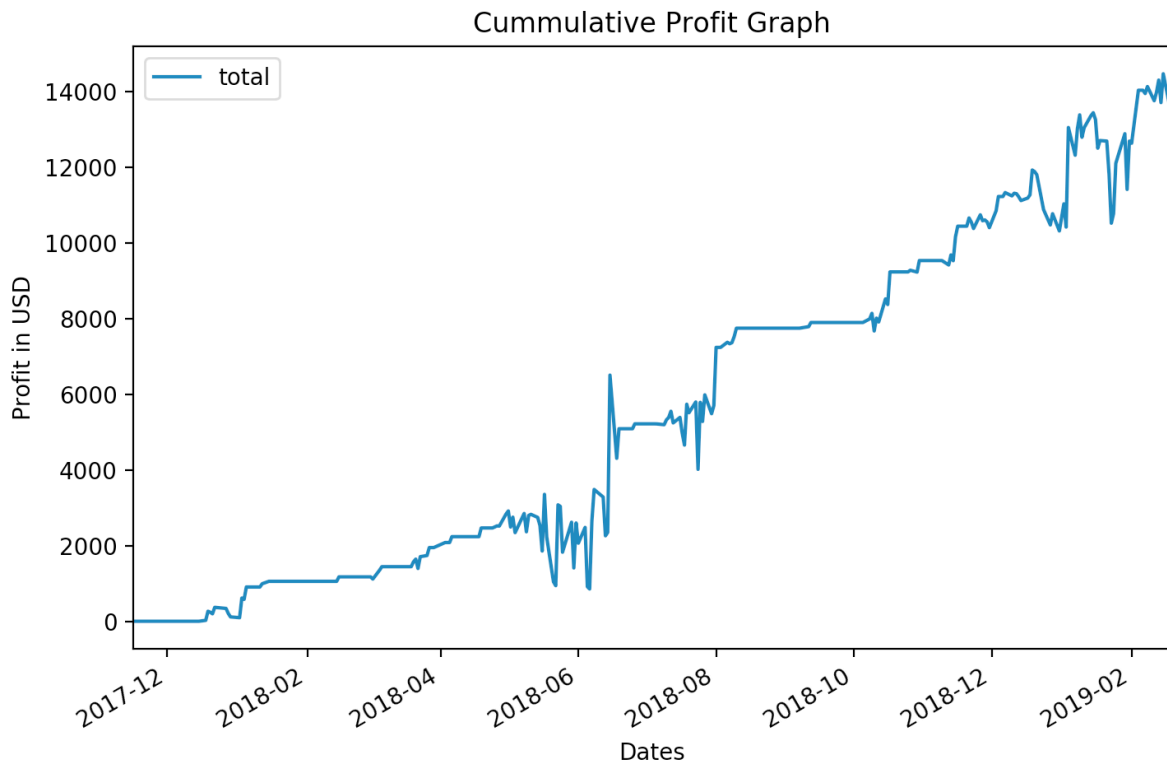


Net Profit: \$16312.2307456

Net Return is: 32.6557257789%

The net profit and return is over a period of roughly 2 years for the training set

Test Set



Net Profit: \$13573.5448883

Net Return is: 15.1472450083%

The net profit and return is over a period of roughly 1 year for the test set

5. Limitations:

- a) The model does not account for transaction cost for trades
- b) Assumption of being able to buy fraction number of stocks to perfectly hedge the position
- c) Because of small dataset, we are only using 10% significance level instead of 5%.

Risks:

- a) Forex risk because the entire model is based on USD
- b) Execution Risk because we need to long and short simultaneously
- c) Australian Market opens before the London Market

6. Future work to improve model:

In order to improve the model further:

- a) Would try to get more data, in order to use a better significance level to be more confident about my results.
- b) Would try to get the dividend information to calculate the adjusted close price instead of the normal close price; and the impact it has on short selling the stock!