FINAL WEEK DOCUMENTATION

WIDS 2024/25

Shubh Tangri

So, this week mainly what we were supposed to do was to apply the trading strategy we had been working on one pair to multiple pairs simultaneously for better and more assured returns, while also analyzing risk mitigation and reliability of the profit/losses obtained from all the pairs used.

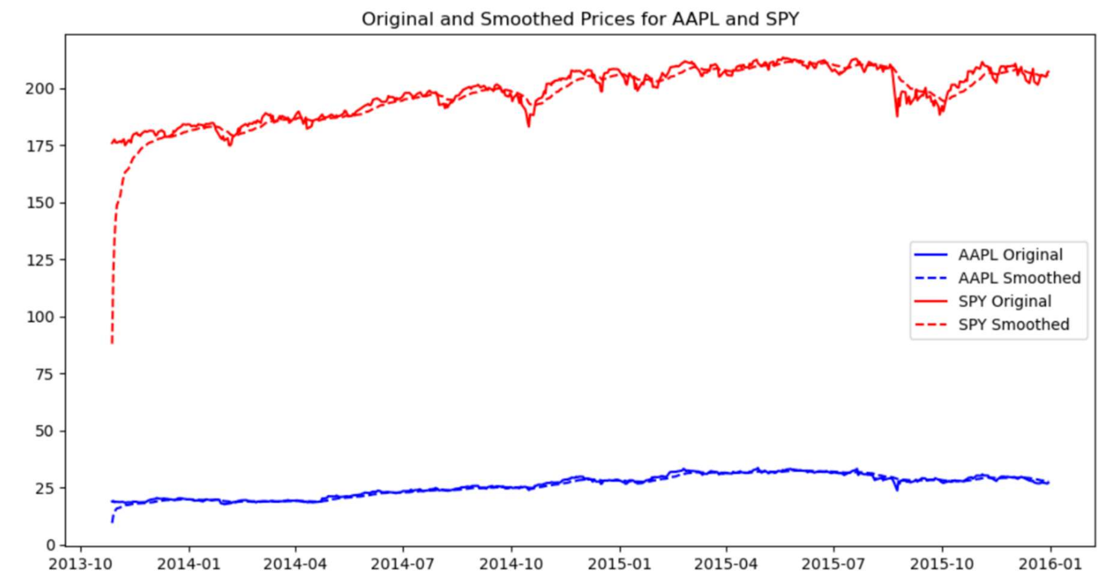
Kicking off, we used our yfinance library to get the data for all the stocks we want. Since we need to find multiple pairs that hold co-integration, we must have a big sample space for our assets. The tickers of the assets used in the code provided are as follows:

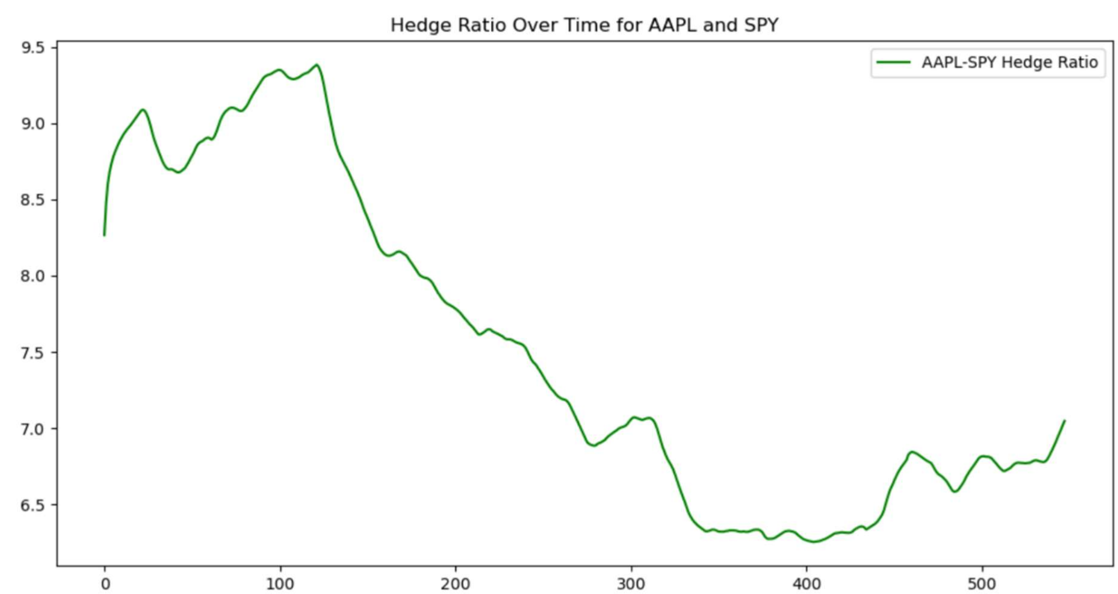
```
import pandas as pd
import pandas_datareader as pdr
from datetime import datetime
import yfinance as yf

def get_historical_Data(tickers):
    """This function returns a pd dataframe with all of the adjusted closing information"""
    data = pd.DataFrame()
    names = list()
    for i in tickers:
        data = pd.concat([data, pd.DataFrame(yf.download(i, start=datetime(2013, 10, 27), end=datetime(2015, 12, 31)).iloc[:,4])], axis = 1)
        names.append(i)
    data.columns = names
    return data

ticks = ["DPZ", "AAPL", "ADBE", "GOOG", "TSLA", "AMD", "GME", "SPY", "NFLX", "BA",
    "WMT", "GS", "XOM", "NKE", "BRK-B", "MSFT", "AMZN", "META", "NVDA", "JNJ",
    "V", "MA", "UNH", "HD", "PG", "JPM", "KO", "PEP", "LLY", "PFE",
    "CVX", "TMO", "COST", "DIS", "ABT", "MRK", "INTC", "CMCSA", "AVGO", "NVO",
    "ORCL", "CRM", "TXN", "QCOM", "UPS", "PM", "DHR", "LIN", "BMY", "RTX",
    "UNP", "AMT", "LOW", "MDT", "T", "VZ", "HON", "NEE", "ELV", "BAC",
    "IBM", "AMD", "SPGI", "PYPL", "AMGN", "SCHW", "AXP", "BKNG", "BLK", "DE",
    "GE", "PLD", "INTU", "ISRG", "MMM", "MO", "LMT", "SYK", "TJX", "ADI",
    "WM", "C", "CVS", "GILD", "CAT", "DUK", "SO", "CCI", "MS", "MDLZ",
    "TJX", "SBUX", "ZTS", "ADP", "F", "PGR", "CB", "CL", "WM", "TRV"] #Name of company (Dominos pizza)
d = get_historical_Data(ticks)
print(d.shape)
# Most Recent Data
d.tail()
```

Next, we loaded the data of all these stocks from 2013 October to 2015 December, and created a heatmap to check for correlation among the stocks. A heatmap of 100 x 100 would quite predictably not be a clear graph, but from the heatmap, one can extract the pairs which have high correlation, for instance those having a correlation greater than 0.85. From these smaller groups of pairs (having correlation>0.85) we can now apply some test to check for co-integration, for example the engle-Granger test, and set a threshold (<0.025) to obtain pairs of stocks having high co-integration. Those pairs of stocks have been mentioned in the attached jupyter notebook.

By now, what we have in our hand is a very strong list of pairs that exhibit high co-integration. We can apply our trading strategy that we have been using in the previous weeks. Basically, using the hedged spread of each pair as the time-series, and check if the z-score values go beyond a threshold or not (to generate the buy and sell signals). Next we check for their hurst components (rejecting those above >0.5), applying KalmanFilterAverage to smoothen the curves, and Kalman Filter Regression, to calculate the corresponding hedge ratios, and then applying the trading strategy.

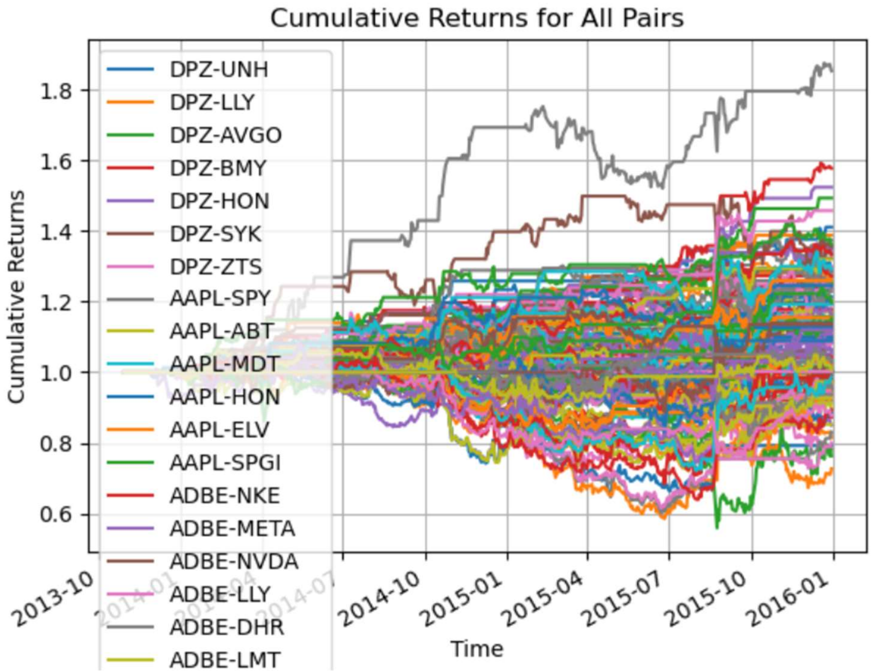
I will show the graphs for the AAPL-SPY pair, as it was the best performing pair among all else :)



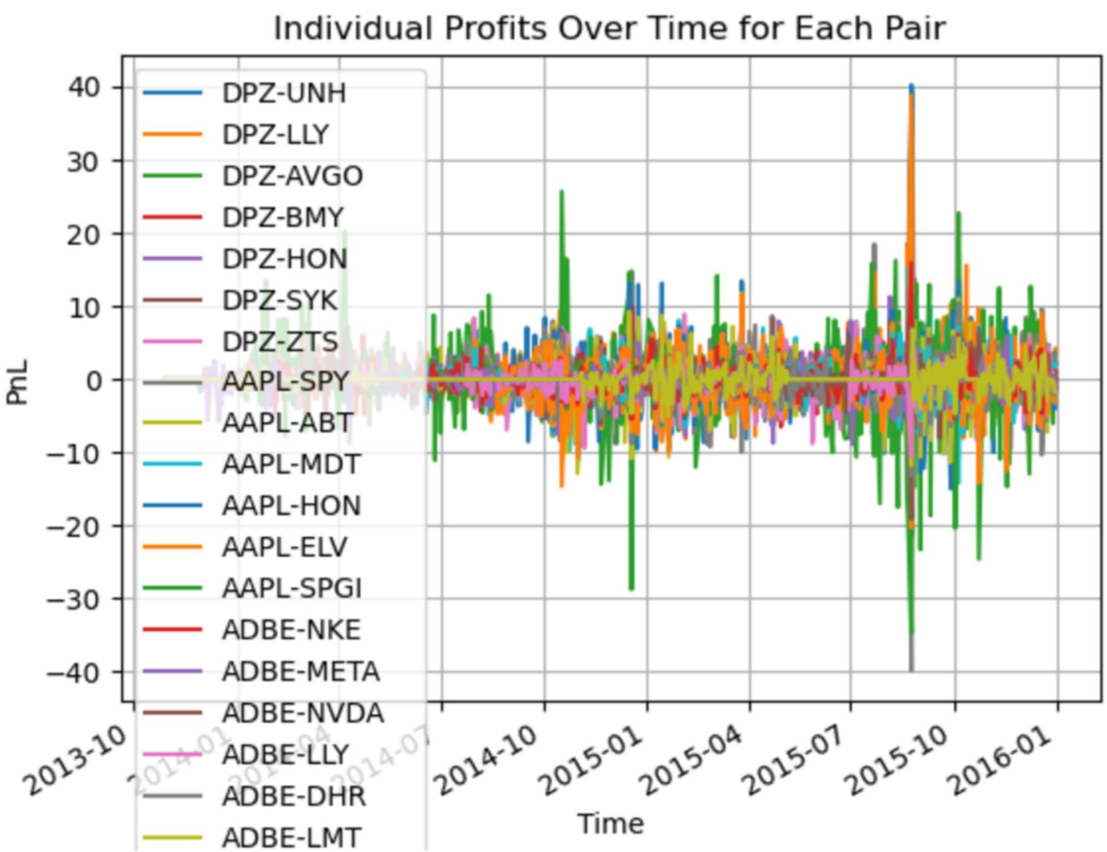


This is how the hedge-ratio over time looks like. The hurst component for the spread of the pair comes out to be a good 0.3974, which makes it eligible for our use. (As a matter of fact all pairs get hurst components less than 0.5 and hence were eligible for our use)

Now, we apply our trading strategy. Calculating the hedge-ratio using the KalmanFilterRegression function, and using it to generate buy and sell signals of the pair if the z-score value goes above/below the set threshold, and calculating the cumulative returns and PnL graphs. For all the pairs used in the code, the graphs looked something as follows:



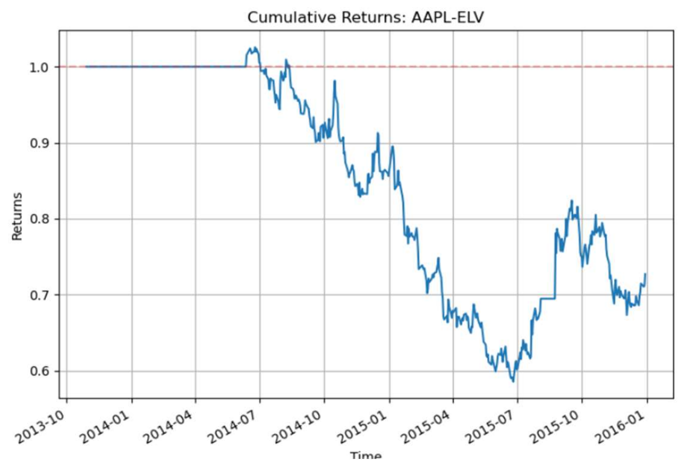
Pretty messy! But, a lot of them seem to be above 1, which is a good omen! To analyse we further calculate the individual PnL obtained from each of the pairs.



Quite clearly, this is also not conclusive. We have good profit bearing stock pairs, but also some ridiculously loss bearing pairs as well, lets analyse them, and find the total profit/loss obtained by applying this strategy on multiple pairs. Here, is how the best and worst performing shares looked like:

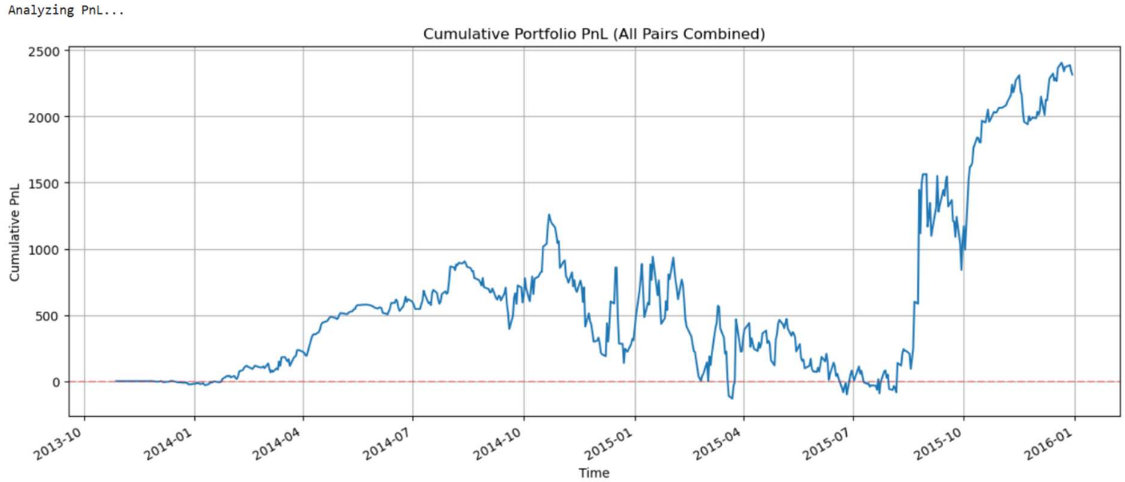
Best Performing: AAPL-SPY

Worst Performing:AAPL-ELV



Here, is what the cumulative profit/loss looked like:

```
Analyzing PnL...
```



A profit bearing portfolio, bearing a profit of above 2300, pretty decent strategy to say the least.

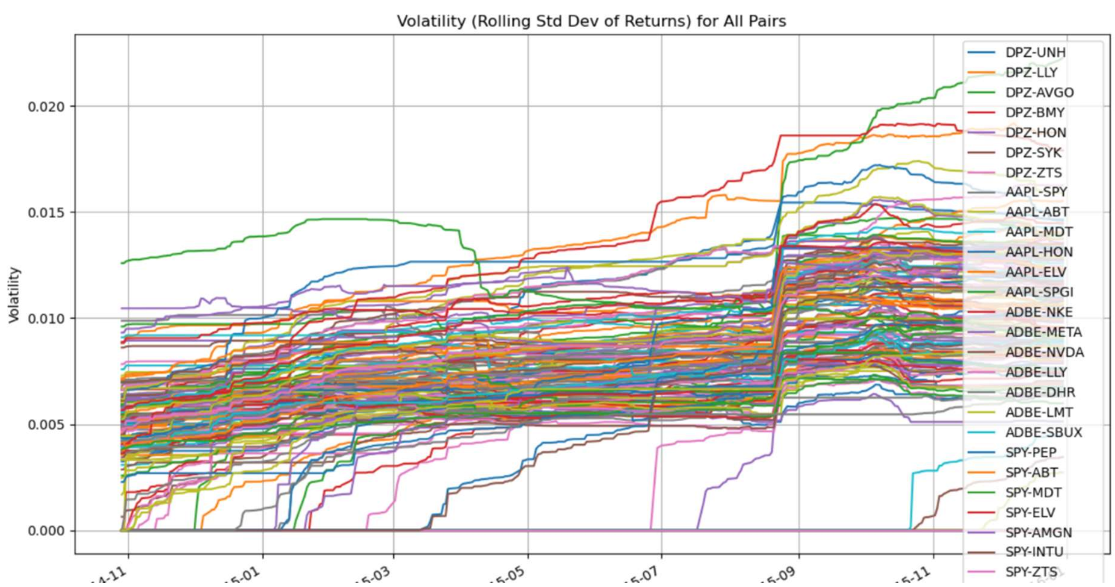
Some other metrics for our portfolio looked like:

```
Portfolio Statistics:
Total Portfolio PnL: 2,315.28
Average PnL per Pair: 13.70
Best Performing Pair: AAPL-SPY (147.74)
Worst Performing Pair: AAPL-ELV (-70.82)
Portfolio Sharpe Ratio: 0.69
Profitable Pairs: 104 out of 169
```

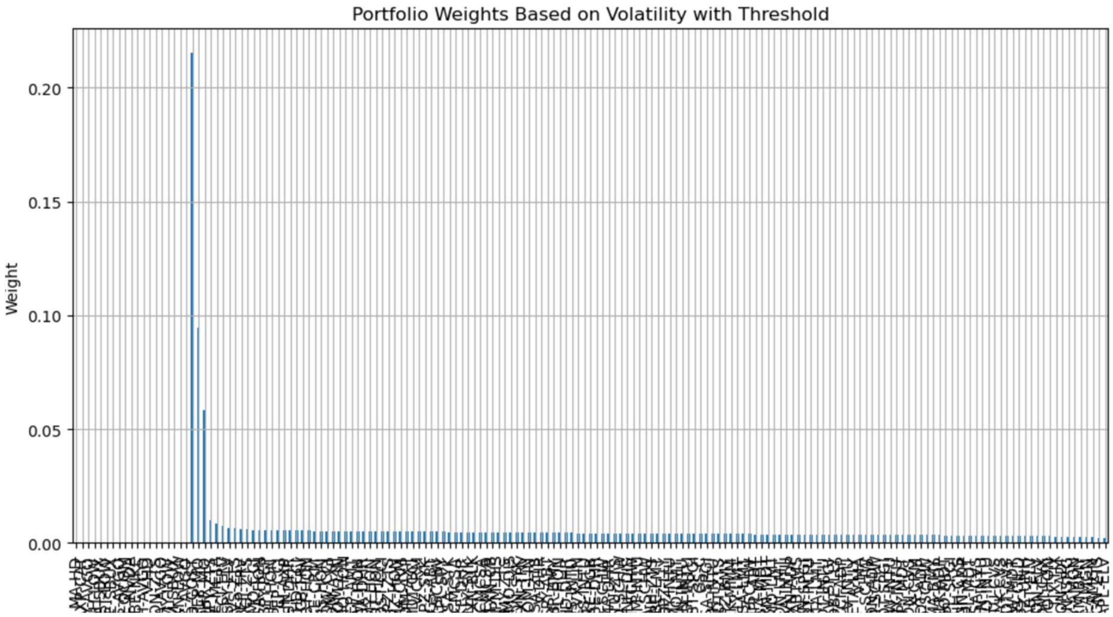
Now, a sharpe ratio of 0.69 is definitely not convincing enough, so I calculated the sharpe ratio for all the profit bearing pairs and loss bearing pairs, to get an idea of which stock pairs are more reliable, and the results obtained were pretty convincing:

```
Profit and Loss Analysis:
==================================================
Total Profits: 3,057.42
Total Losses: -742.13

Sharpe Ratio (Profits): 1.38
Sharpe Ratio (Losses): -0.58
```

A positive ratio and that to above 1 for profit bearing stocks is a good sign, and we have to some extent mitigated the risk. In the notebook attached, I have arranged all the stock pairs in the descending order of their returns, which can be used to give weights to each of the stock pairs to amplify our returns (I believe). However, we also have to be certain about, whether or not the pairs have incredibly high volatility or not, to assure risk-mitigation. So, here is the graph for volatility:



On the basis of volatility of returns, we can give weights to all the stocks in inverse proportion to their volatility, meaning the lesser volatile return bearing pairs, will have a higher weight in our portfolio. This can assure stronger, more reliable returns if not “High” returns. The consequent bar graph for the weights is as follows:



Thus, to conclude, the applied strategy on multiple pairs of stocks, quite evidently yield a better more reliable return as opposed to the individual pairs. Out of the 169 used pairs, over 100 of them were profit bearing, and we can easily weight them on the basis of two parameters: how much profit have they borne in the past years? And how volatile have their returns been over time. Without the weights the obtained profit was decent enough, but with the weights the profit might rise/decline but it will surely have a better over all sharpe ratio and reliability. It is quite incredible how simple concepts of stationarity of time series and mean reversion can do wonders in the world of finance. A very intriguing project and a very good learning experience over the past few weeks. Thank you so much for the opportunity!