Project A

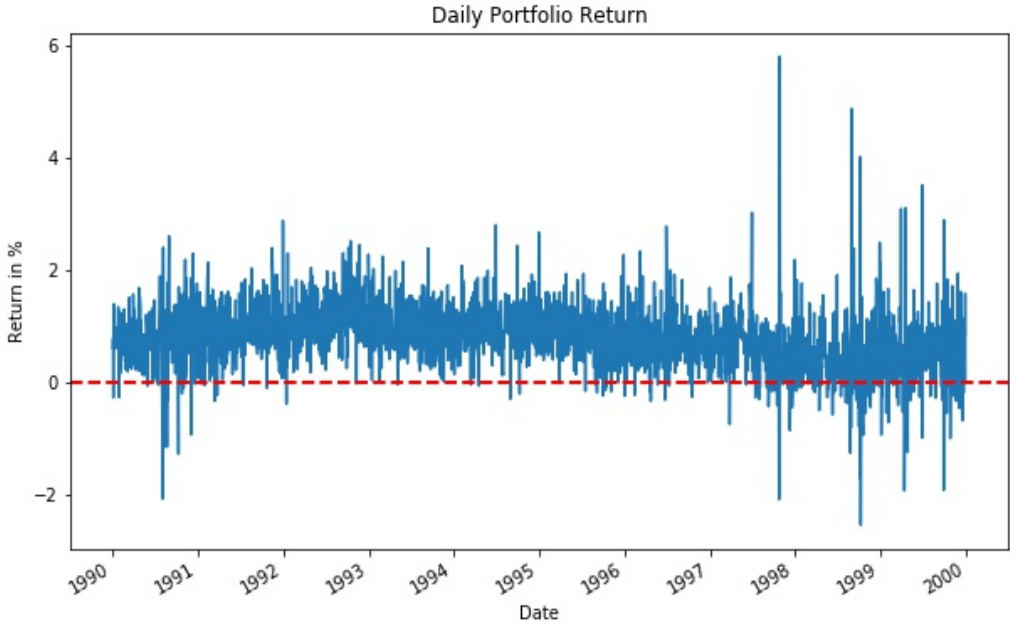
Raj Kumar Anand
rajanand@mit.edu

September 30, 2020

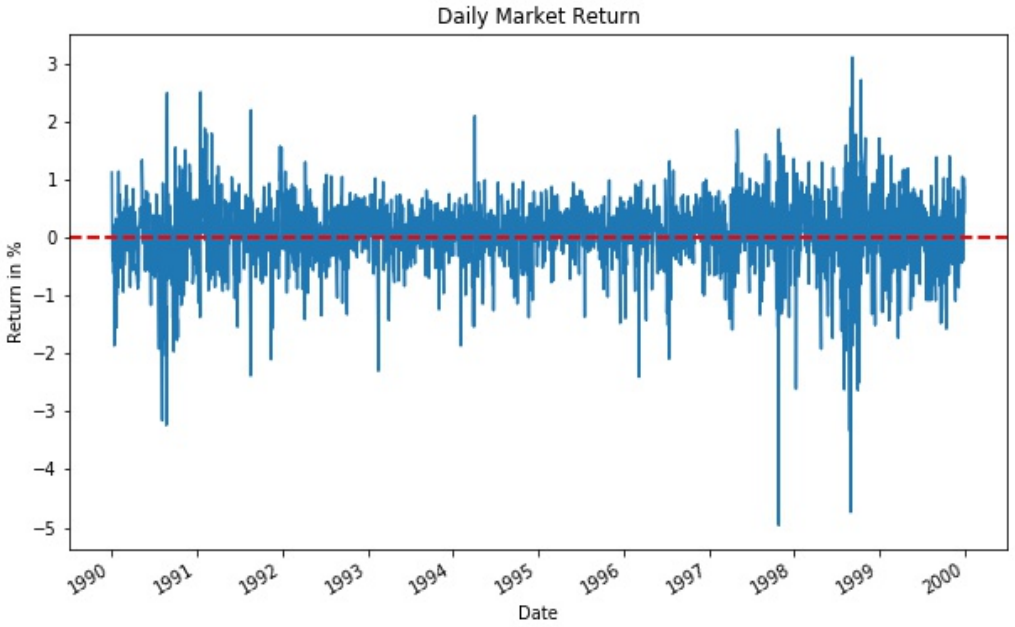
Exercise 1

The strategy is created using data provided to us on SQL database ('crispy'). The database contained three tables having log-returns for different securities on different dates, along with other information like ticker, IDs and security names. Using SQL Server Management, data is structured in a one table, satisfying all the criterion mentioned in the problem statement and then exported to flat files.

(a) After creating the required strategy from historical data, following are the returns for Portfolio and for Market.



As evident in above figure, daily returns of the portfolio are usually above 0, suggesting that the strategy is performing positively most of the time. Whereas, the market return shown below is varying around 0.



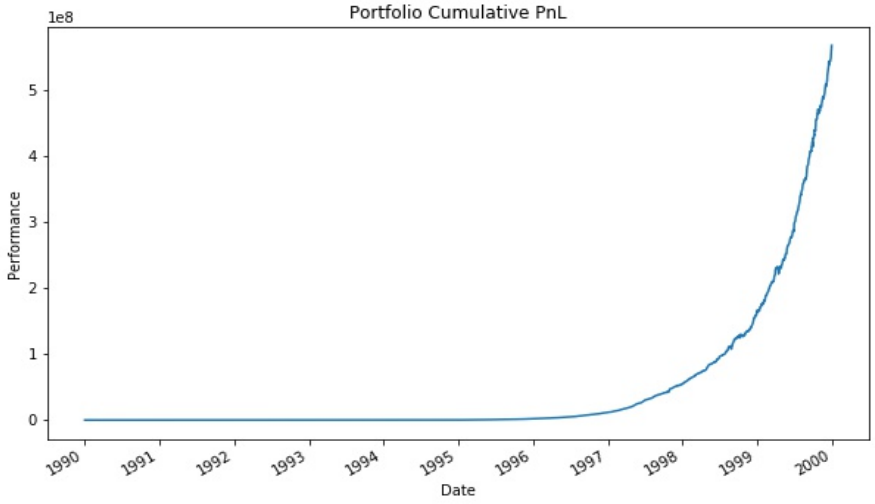
(1)

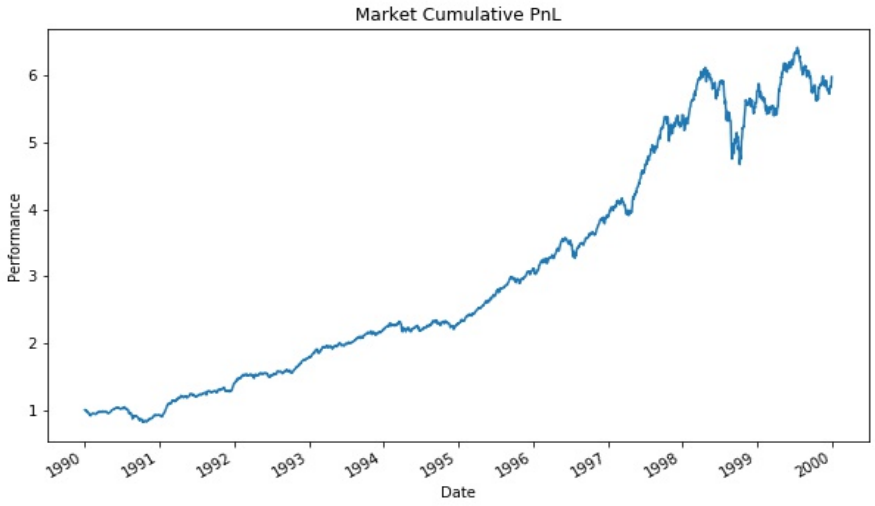
While looking at return series can give a fair ideas about the performance, but it does not tell us how good the strategy does exactly. For getting a more detailed picture of performance, we need to calculate basic statistics. These statistics along with cumulative PnL is given in section (b).

(b) Annualized mean return for Strategy and Market is given in the table below.

	Strategy	Market
Annualized Return	201.62%	18.29 %
Volatility	9.41 %	9.64 %
Sharpe Ratio	21.43	1.90

Further, to complement the above statistics and to gauge the overall performance, we can look at cumulative PnL which is equivalent of finding the value of 1 dollar, invested at the start of the strategy over the entire testing period. The cumulative performance of strategy and market is shown in the charts below.





(c) For checking stationarity of return series, Augumented-Dickey Fuller test [1] is been used which tests for unit root. The null hypothesis (H0) states that series is non-stationary. Following table summarizes the result of test with threshold value of statistics for different confidence interval.

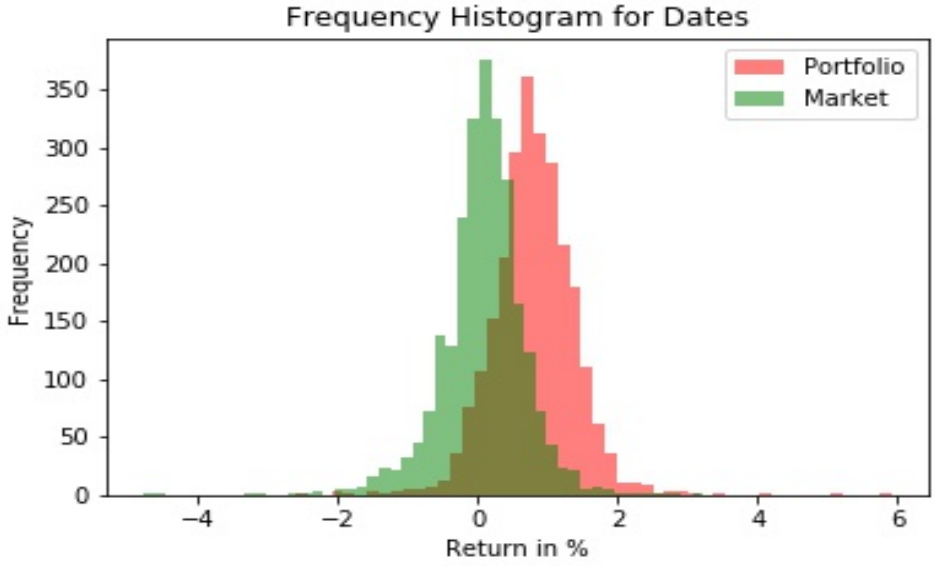
	Strategy	Market
ADF Statistics	-3.758783	-10.220757
p-value	0.003358	0.00000
Result	Stationary	Stationary

Confidence Interval	99%	95 %	90 %
Threshold ADF Statistics	-3.433	-2.863	-2.567

Both series seem to be stationary, p-value ≤ 0.05: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

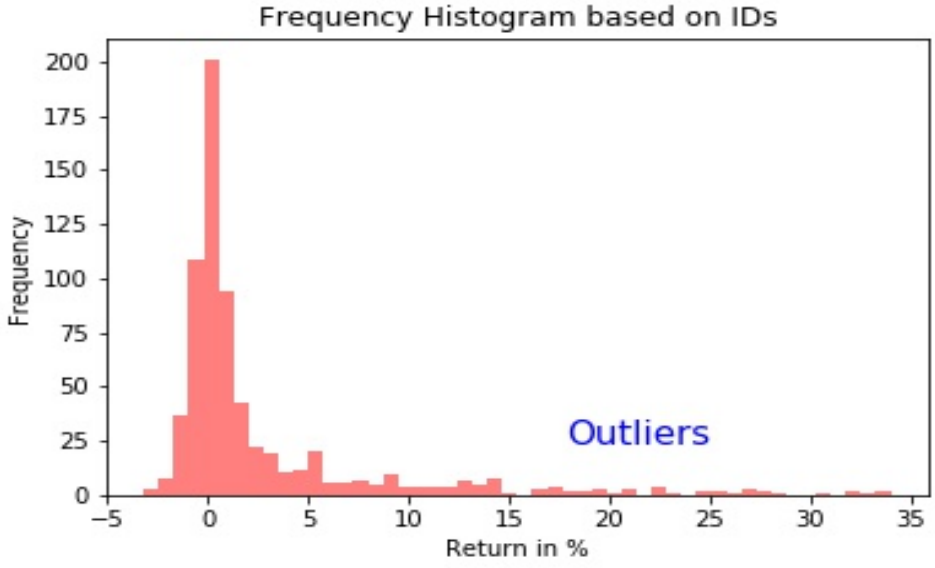
(d) Outliers can be detected based on different criterion like events based on dates, event based on securities. These can further be divided into finding outliers based on overall contribution to the portfolio or number of positions.

Checking Returns based on dates In order to check any outliers based on dates, we can sum the total returns over all the securities for each date and check the distribution of that return. This is shown in figure below. As evident from the plot, the distribution for both market and strategy is seems to be distributed normally. The center of Porfolio is towards the right, which is inline with the excess performance that we saw earlier.

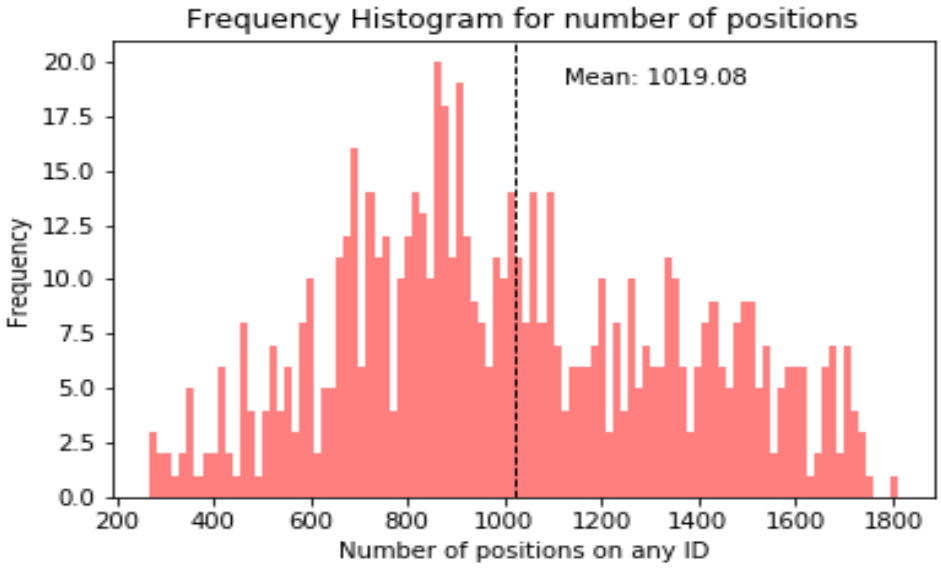


Checking Returns based on dates In order to check any outliers based on Securities, we can sum the total returns over all the dates for each security and check the distribution of that return. This is shown in figure below. As evident from the plot below, there are many securities which have disproportionately high positive contribution to the strategy. There are even some security that are adding a total of 35% of returns.

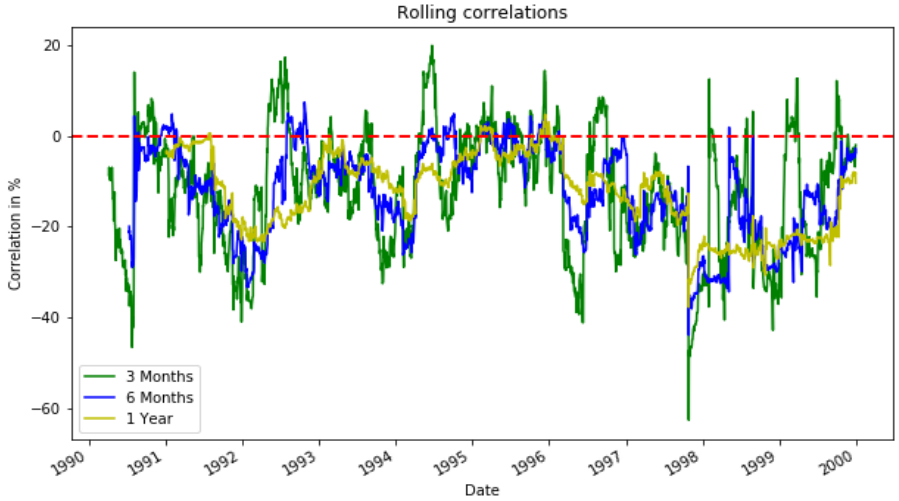
As shown in part (b), portfolio return is greater than 200% while the market return is only around 18%. This alludes to the impact of these outliers can have on the strategy performance.



Checking number of times a security held a position in the strategy In order to check the number of times each security contributed to the performance of strategy, we need sum the absolute values of positions (1,-1 or 0) over all dates.
As observed in the plot below, out of 2528 trading days in our sample, some of the securities are included in the portfolio for 1800 times. This means that our portfolio is very concentrated and it's performance will be affected by a small number of securities.



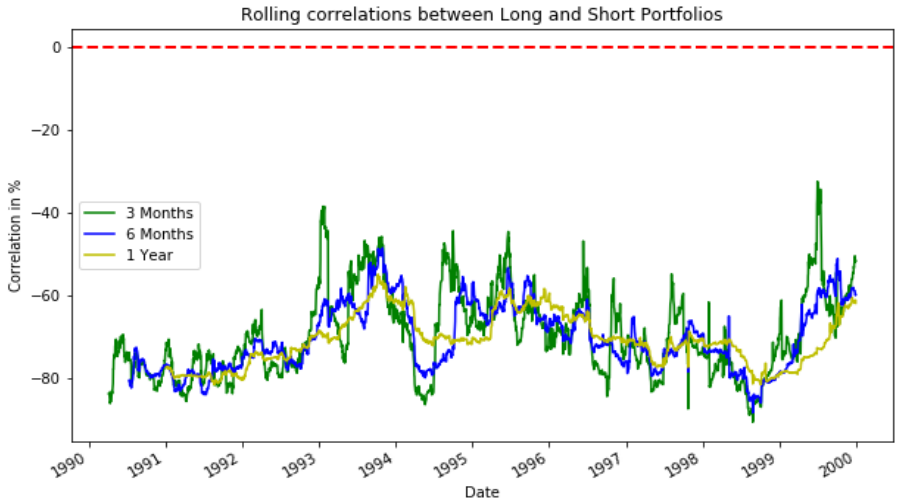
(e) Historical average correlation between portfolio return and market return is -11.95%. In addition of finding correlation, correlation can also be viewed at different times based on different time windows. Time series of correlation is plotted for 3 different time windows.



As evident from the plot, historically the strategy is negatively correlated with the market for all the three time windows.
This shows that the strategy is just dollar-neutral and not market neutral.

(f) A Long portfolio is constructed by using weights that were positive while a short portfolio is constructed by using only the negative weights in our original construction.

Historical average correlation between Long Portfolio and Short Portfolio is -68.57%. In addition of finding correlation, correlation can also be viewed at different times based on different time windows. Time series of correlation is plotted for 3 different time windows.



As evident from the plot, historically long and short portfolios are negatively correlated but if the strategy were to remain market-neutral, these strategies should have a correlation of -100%. But there correlation is not as negative as they should have been.

(g) The performance of portfolio constructed above is based on various simplifying assumptions due to which the actual return (when applied in actual financial markets) can be very different. Some of these assumptions are listed below.

- **No transaction cost** The current strategy is changing its exposure everyday based on the signal without considering transaction cost which would be applicable in real setting. In case transaction cost were present, then it would be optimal to rebalance the portfolio on definite interval rather than daily.
- **No market impact** The strategy also assumes that while rebalancing the portfolio daily, there is no impact in the market. This condition may be true only if all the securities are very liquid. Otherwise, while rebalancing the portfolio, market prices may move against us.

Due to the above simplifying assumptions, portfolio return is not very realistic.

Data Issues During the construction phase of this strategy, there were some data issues that can have a lot of impact on the performance of the portfolio.These issues and their possible solutions are discussed in the bullets below.

- There were around 250 days when a large number of securities had 0 returns. One such date is 1990-01-15 when 148 securities have 0 return. This creates a problem when bucketing the signal (based on these returns) into different bins. Ideally, each quantile should be $\frac{690}{5} = 138$ length, but if we have quantile with fixed lengths then 0 return for one security will be treated differently than 0 for another security, which is not fair. So, in order to account for these dates (when there are large number of securities with same return), I have dynamically defined the size of quantiles on each dates. These quantiles

are usually of length 138 (as they should be), but on special dates, the size changes to fairly accommodate all the returns.
This dynamic approaches also enables us to be market neutral (sum of weights equalling to 0) because the weight for each security will be divided by number of security in the quantile, there by allocating lower weights to quantiles with greater number of securities.

- Some of the ticker are not unique. So creating a dataframe of returns with dates as the row index and ticker as the column heading will give wrong result. In order to avoid this pitfall, I have used "id" as a unique identifier. The "id" field in crispy and crispy04 is inherited from CRSP's assignment of its own unique identifiers to traded securities. They also maintain a history of ticker changes (e.g., so that you can answer what company had which ticker on a specific date) and, more importantly, corporate actions involving mergers, spinoffs, etc.

Exercise 2

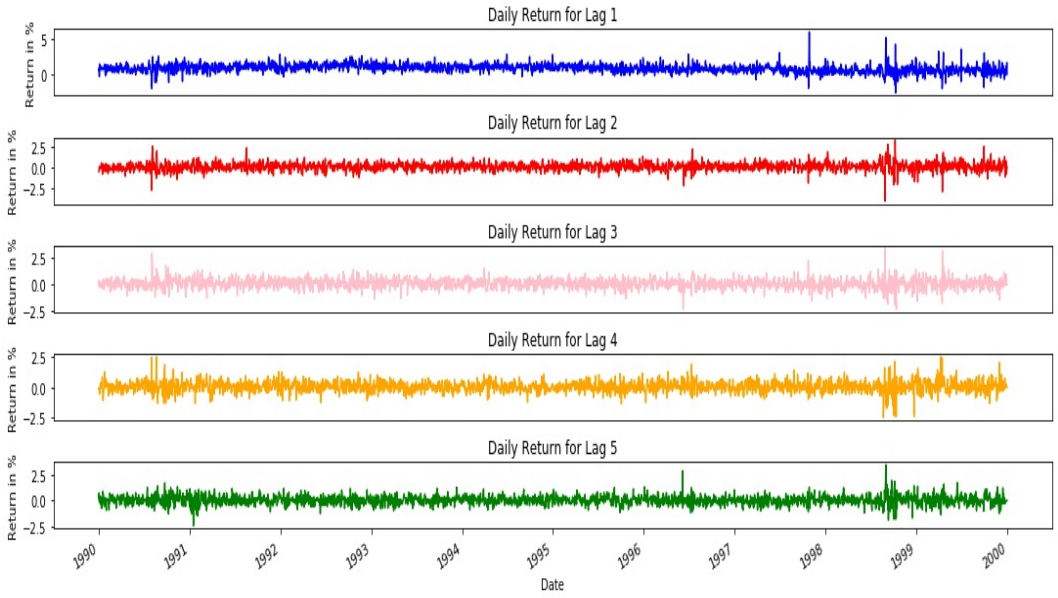
The original strategy is generalized by varying the lag k. The generalized signal is defined as follows.

$$s_i^{(k)}(t) = -[R_i(t-k+1) - \bar{R}(t-k+1)]$$

(a) Performance of the portfolio for different values of lag is summarized in table below.

	1 lag	2 lag	3 lag	4 lag	5 lag
Annualized Return	201.62 %	20.09 %	17.46 %	13.55 %	10.20 %
Volatility	9.41 %	7.82 %	7.24 %	7.34 %	7.16 %
Sharpe Ratio	21.43	2.57	2.41	1.84	1.43

The performance of different versions of strategies is shown in the plot below.



Following observations can be made from the statistics and the plots.

- Annualized return falls sharply as the lag value is changed from 1, suggesting that performance of this strategy is highly susceptible to change in the lag value.

- The performance of the strategy also sees a steady fall as the value of lag is increased, indicating that the contrarian view on which the strategy is based becomes weak as the lag value is increased.
- From the graph, it is evident that all the five versions of the strategy have very similar return profile. But there are differences in the return, unique for each value of lag. For example, around 1998, Lag 1 has large positive return which is absent in other versions.

In terms of choosing an optimal strategy, as seen from above, strategy with lag 1 is very unstable, as a slight change in value of lag changed its performance drastically. But from lag value of 2 and onwards, the performance has remained more stable. So, if an investor were to choose a robust strategy with sound performance, Strategy with lag of 2 is the optimal strategy as it's performance is stable and also highest compared to other versions (except lag =1).

(b) For Question 2 (b) and other subparts, flat files containing the following variables (columns) has been uploaded to **ProjectB** of database **upload**.

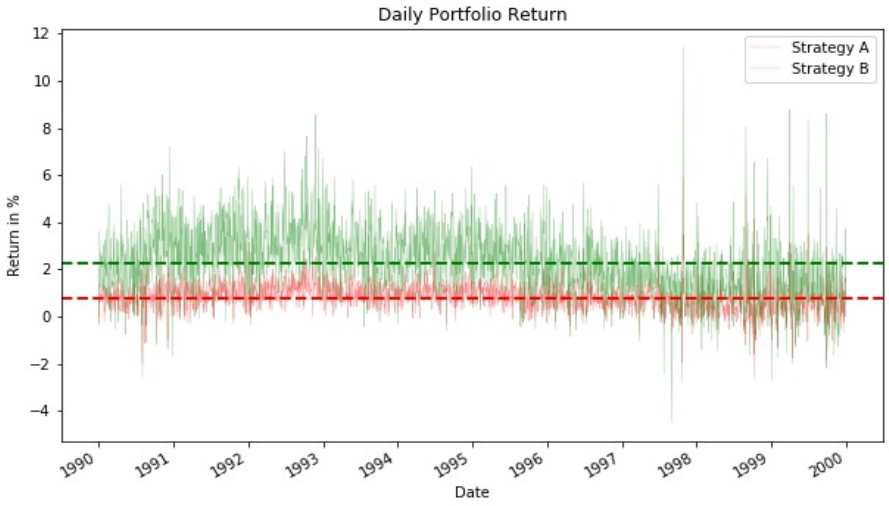
- Column 1: Portfolio Manager ID pid (926773554)
- Column 2: Date d
- Column 3: Stock ID id
- Column 4: Time lag k
- Column 5: Portfolio weight w
- Column 6: Version ID vid (Set equal to 0 for all rows)

Exercise 3

One of the issues that pointed out in Question(1) is the concentration of some securities. In order to avoid this concentration, one of the possible technique can be reducing the number of total positions in the market.
To continue this exercise, instead of taking positions in top and bottom quintile, we will simulate the performance taking position in top and bottom 5 percentile. Reducing the number of positions may also result in reduced annualized return, because we would be have less long or short exposure to securities which should go up or down, based on our investment rationale. But we are more concerned with risk-adjusted return so we will be comparing Sharpe Ratios.

Name	Description
Strategy A	Going Long top Quintile and going short bottom quintile
Strategy B	Going Long top 5th percentile and going short bottom 5th percentile

(a) The returns of both of the strategy is compared in the figure below. As evident from the below plot, average return (marked with dotted lines) for Strategy B is higher than our previous version of the strategy. This overperformance is in contrast with our original expectation when we expected lower return due to lesser number of positions (overall exposure). We need to to delve down further to find the reason for this enhanced performance.



(b)

	Strategy B	Strategy A	Market
Annualized Return	574.62%	201.62%	18.29 %
Volatility	23.14%	9.41 %	9.64 %
Sharpe Ratio	24.83	21.43	1.90

(c) Checking the stationarity of time series over time using Augmented-Dickey Fuller test.

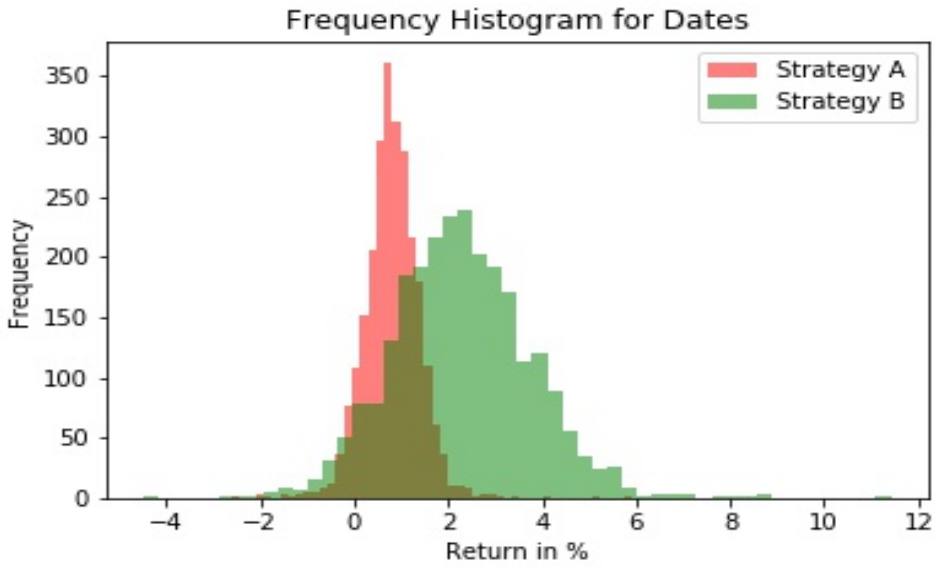
For checking stationarity of return series, Augumented-Dickey Fuller test [1] is been used which tests for unit root. The null hypothesis (H0) states that series is non-stationary. Following table summarizes the result of test with threshold value of statistics for different confidence interval.

	Strategy B	Strategy A	Market
ADF Statistics	-3.200492	-3.758783	-10.220757
p-value	0.019956	0.003358	0.00000
Result	Non-stationary	Stationary	Stationary

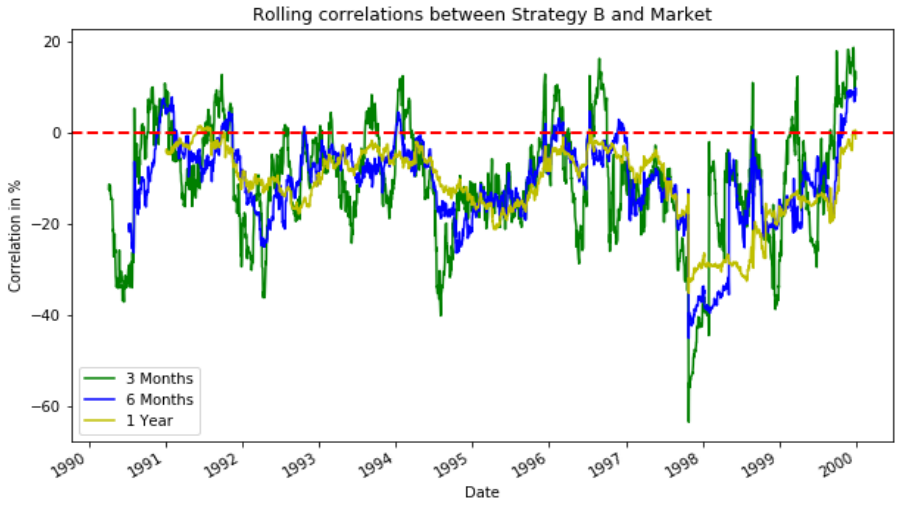
Confidence Interval	99%	95 %	90 %
Threshold ADF Statistics	-3.433	-2.863	-2.567

Return series for strategy B is not stationary with 99% confidence interval.

(d) The figure below plots the distributions of the return for both the strategies. As seen in the plot, the distribution of the strategy B is more wider than the strategy A, suggesting a lower kurtosis for strategy A.

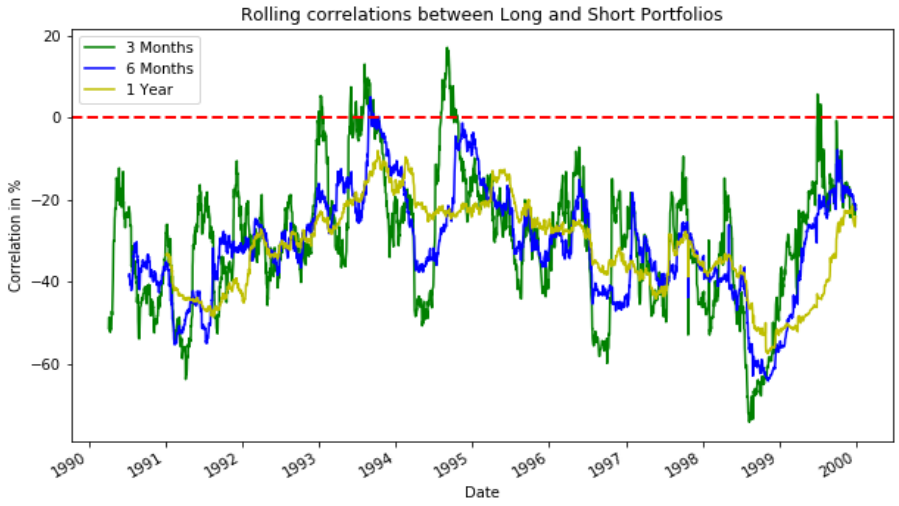


(e) Historical average correlation between strategy B and market return is −8.26%. In addition of finding correlation, correlation can also be viewed at different times based on different time windows. Time series of correlation is plotted for 3 different time windows.



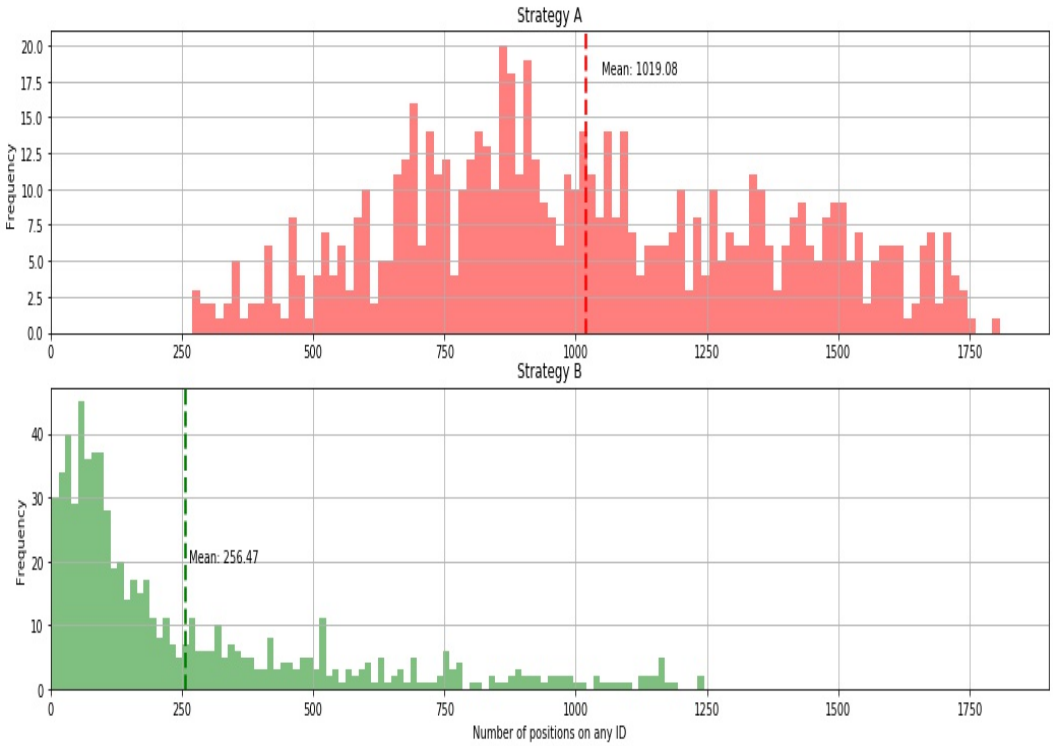
As evident from the plot, historically the strategy is negatively correlated with the market for all the three time windows.
This shows that the strategy is just dollar-neutral and not market neutral.

(f) A Long portfolio is constructed by using weights that were positive while a short portfolio is constructed by using only the negative weights in our original construction.
Historical average correlation between Long Portfolio and Short Portfolio is −26.19%. In addition of finding correlation, correlation can also be viewed at different times based on different time windows. Time series of correlation is plotted for 3 different time windows.



As evident from the plot, historically long and short portfolios are negatively correlated but if the strategy were to remain market-neutral, these strategies should have a correlation of -100%. But there correlation is not as negative as they should have been.

(g) Our original rationale for constructing this strategy was to reduce the concentration of the portfolio that we were witnessing in our previous Strategy A. The figure below plots the number of positions taken by each ID (on the x axis) in the portfolio over our testing period.



The above plot suggests that our original strategy had an average of 1019.05 positions, with as high as 1800 positions taken by some IDs. In strategy B, average number of positions is only 256.47. This suggests that our strategy A has less concentrations than strategy A. But this is also due to lesser number of securities invested at any given time in strategy B.

References

[1] Augmented Dickey-Fuller Test,
https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/