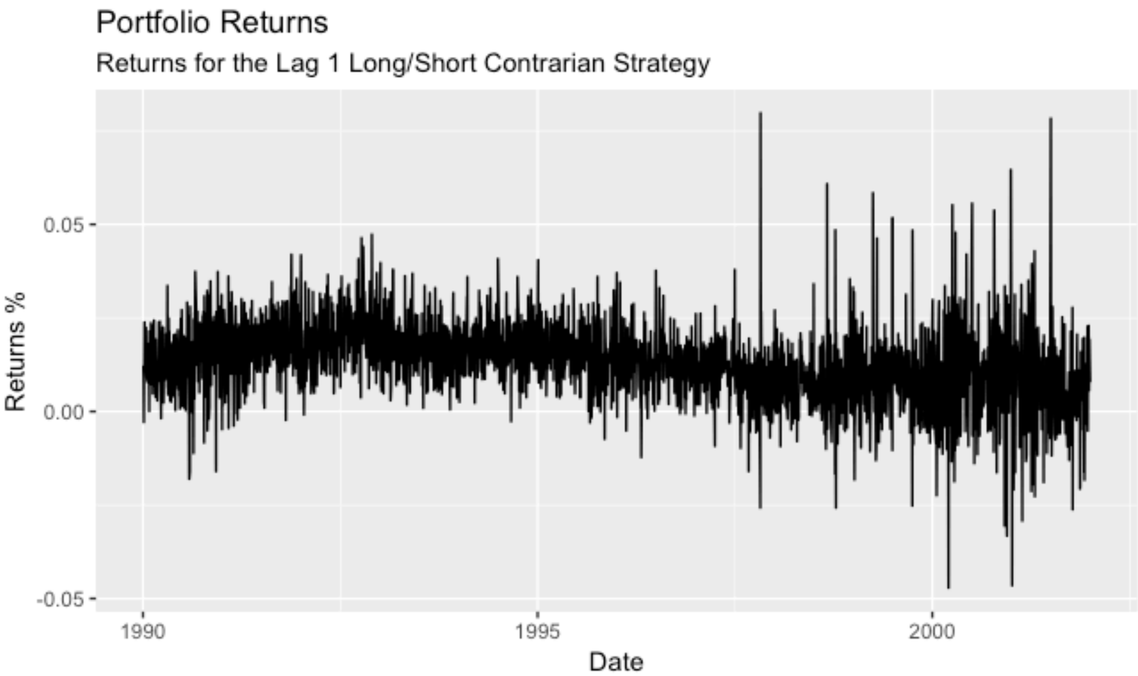
15.458 Financial Data Science and Computing

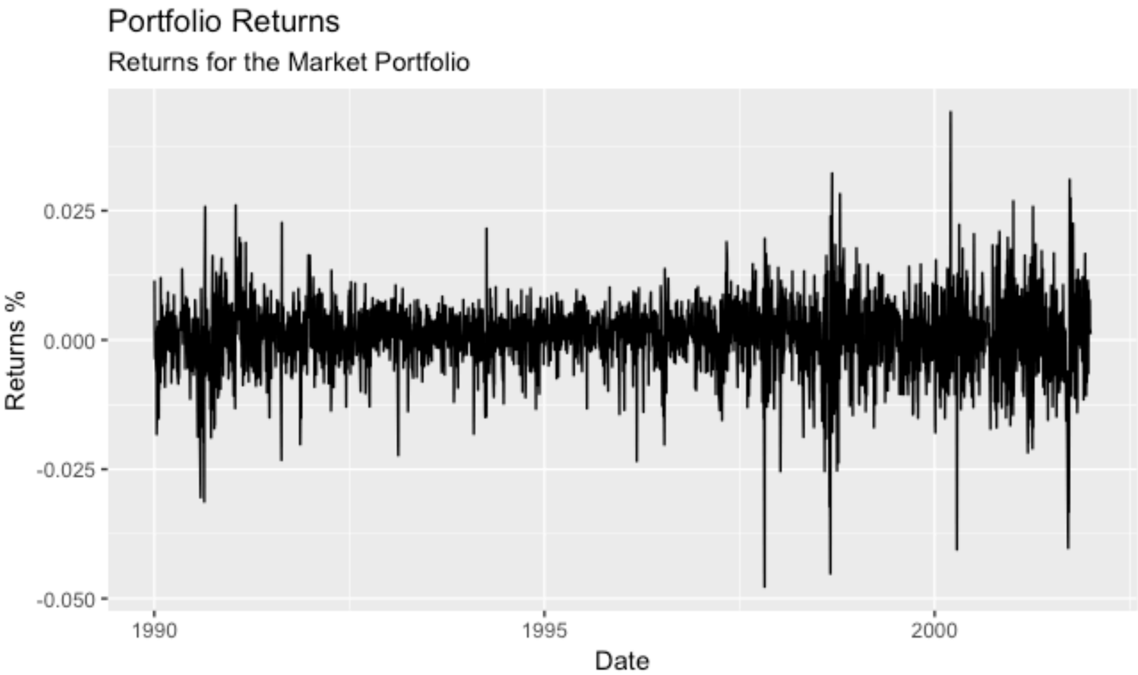
Devin Connolly (929315784)

30th of September 2019

Question 1 – Strategy Simulation

(a) The returns of the long/short contrarian strategy was constructed and back-tested. The returns of both the contrarian strategy and daily market returns are shown below:





(b) The summary statistics (mean, volatility, Sharpe ratio) for the portfolio were as follows:

Annualised Mean	331.78%
Annualised Standard Deviation	16.36%
Sharpe Ratio	20.28

For the market returns:

Annualised Mean	17.72%
Annualised Standard Deviation	10.63%
Sharpe Ratio	1.67

(c) The portfolio returns, volatility, and Sharpe ratio are not consistent over time. A simply way to test this is to split the data into two equal halves and evaluate the metrics for each:

First Half

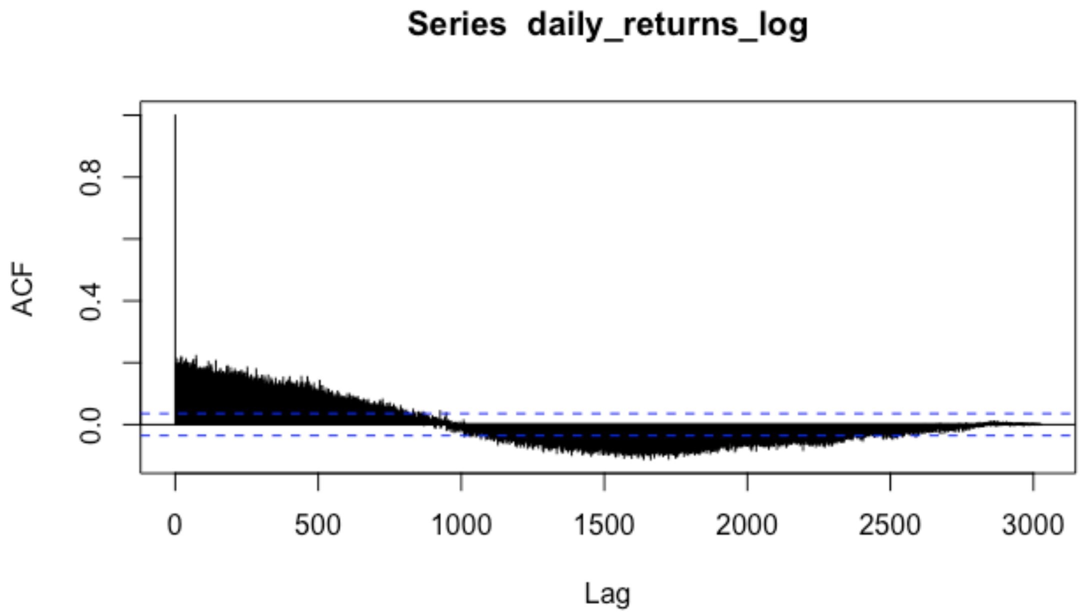
Annualised Mean	431.47%
Annualised Standard Deviation	12.65%
Sharpe Ratio	34.1

Second Half

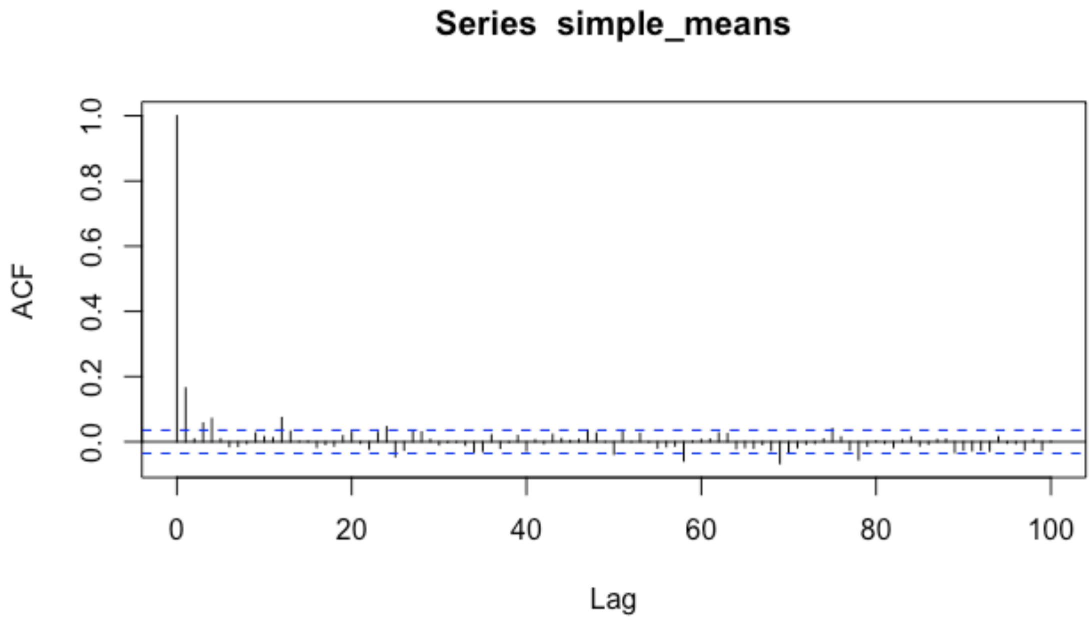
Annualised Mean	232.24%
Annualised Standard Deviation	17.21%
Sharpe Ratio	13.49

The first half of the time horizon used significantly outperformed the second half, illustrating the performance metrics are not consistent over time. This could be further tested by breaking the time period into even smaller intervals.

In order to check if the portfolio returns were stationary in time, the Ljung-Box test and plotted the ACF. The Ljung-Box test gave a p-value of 2.2e-16. As this is greater than 0.05, we can conclude that the residuals are not independent. Thus, the series is not stationary. The ACF for the time series is as follows:

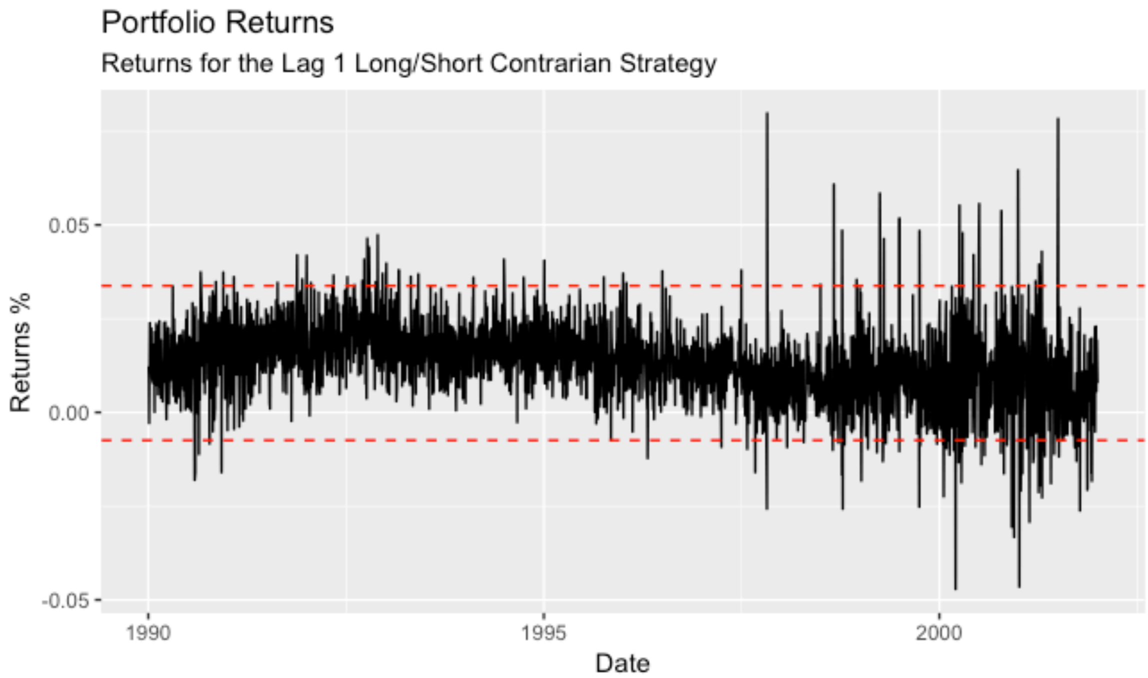


This shows clear correlation in time, indicating that the time series is not stationary. Carrying out similar tests for the market returns yields a Ljung-Box p-value of 2.2e-16 again, but the ACF is plotted as follows (for lag 100 only):



This indicates that there is strong correlation with lag 1, meaning that the market returns are also non-stationary (albeit to a much lesser extent that the portfolio).

(d) In order to begin looking at outliers, I first found the maximum and minimum return over the period. The highest daily return of my strategy was 7.98%, while the lowest daily return was -4.71%. Plotting the daily returns of the portfolio with horizontal lines for plus or minus two standard deviations gives:



When these outliers are removed, a sample of 2898 days. Calculating the performance statistics for this portfolio yields:

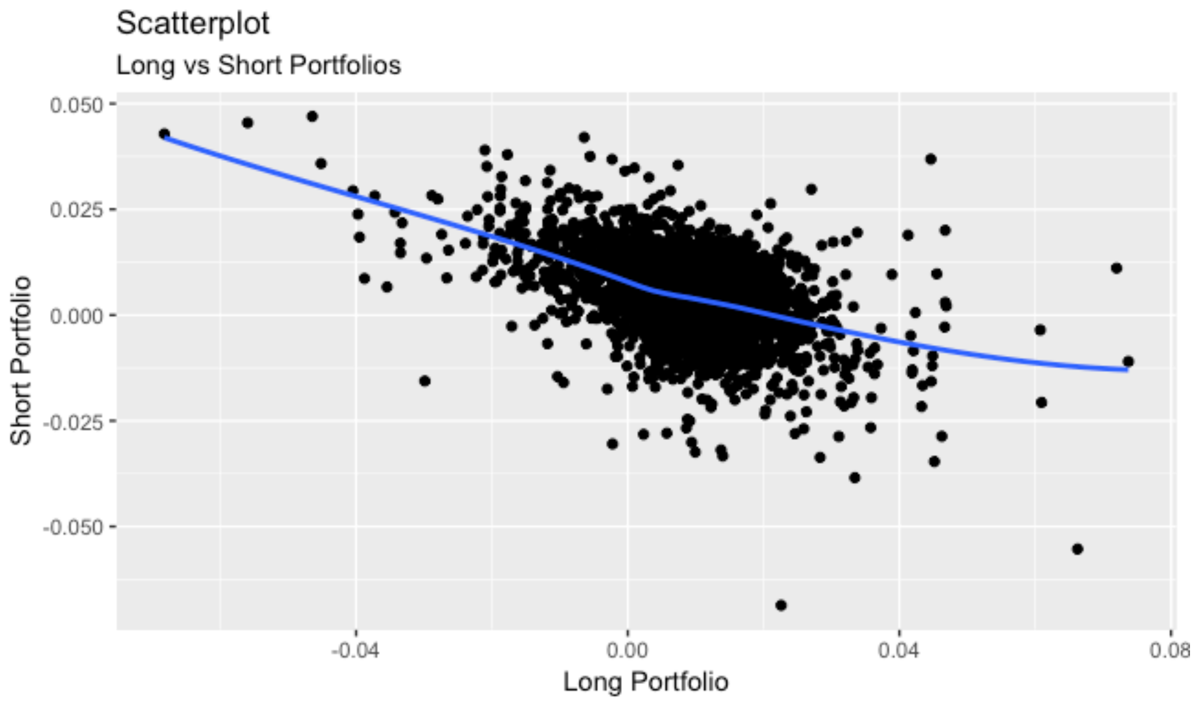
Annualised Mean	335.56%
Annualised Standard Deviation	13.35%
Sharpe Ratio	25.14

It is interesting to note that the portfolio performance improved as a result of removing both positive and negative outliers. This means that outliers have a significant negative effect on overall portfolio returns.

Then, analysing the overall returns by stock, I calculated the maximum return contributed by security to be 77.71% (ID = 66546) and the minimum return contributed was -4.89% (ID = 21338). The fact that the lowest return contributed by an individual security is only -4.89% over the period indicates the strategy is operating effectively. However, the overall portfolio return seems to be reliant on a number of high-performing individual stocks. These stocks (66546 for example) appear well-suited to this contrarian strategy.

(e) The correlation between the portfolio returns and the market returns is 0.1214. The contrarian strategy portfolio is completely dollar neutral, but is not quite market neutral. To be completely market neutral, the portfolio would have to have 0 correlation with the market. In this case, the portfolio does benefit slightly from upward market movements.

(f) The correlation between the long and short sub-portfolios is -0.4675, meaning that the long portfolio performing well is accompanied by the short portfolio doing poorly. The correlation can be illustrated on a scatterplot (using loess to fit a demonstrative line):



(g) This simulation is not very realistic. There are a couple of main reasons for this:

1) The portfolio is essentially allowed a daily turnover of 200% (close out every position and re-open entirely new positions) every day. This can be both difficult and very expense to do in practice (illiquid stocks, market impact).
2) There are no transaction costs included.
3) We are always using the market price to buy and sell. In real life, there will be bid-ask spreads on security trades that will cut into profit margins.
4) In this strategy, instantaneous trades are assumed. This is not realistic, as trades cannot be made instantaneously outside of open-market hours (overnight). Furthermore, there may be slippage in real implementation (trade price may differ from price at time of execution).

Assuming there are no transaction costs and no market impact, the simulation will still be flawed, primarily by the bid-ask spread. It is shown to have worked throughout a back-test, but whether or not it will work on real, live market data in the future is hard to speculate. While there certainly could be returns to be made using a contrarian strategy, now that it is well-known and commonplace, it would be impossible to generate returns to the extent as the returns featured here.

Data issues could also play a factor in the real implementation of this strategy. The most noticeable causes for concern are:

1) The weight calculations are sensitive to live data feeds. It is important that these returns are accurate and complete. It is unrealistic to cover the entire market, so a subset of stocks (a universe, in our case) needs to be established for which weights can be applied to.
2) In certain situations, data even from the most reliable of vendors (Bloomberg) can have back-filled data, or incomplete data that is filled with other data fields (OLHC all being equal, for example). Furthermore, it is imperative that only valid trading days be included in the data feed to the strategy.

Question 2 – A Family of Strategies

Generalising the strategy across five different lags gives the following results:

	1	2	3	4	5
Annualised Mean	335.56%	36.78%	23.17%	17.14%	9.26%
Annualised Standard Deviation	13.35%	13.62%	13.21%	13.11%	13.08%
Sharpe Ratio	25.14	2.7	1.75	1.31	0.71

There is clear downward trend as lag increases. The volatility of the portfolios is relatively constant, but returns and the Sharpe ratio decrease dramatically. The returns of the portfolio decreases exponentially. The optimal value for the lag is then k = 1, the original portfolio. This offers exceptionally high annual returns.

I would argue that the reason the lag 1 portfolio performs the best is because a contrarian strategy aims to capitalise from over and under reactions. This table thus shows that these reactions tend to be corrected (for the most part) within one trading day.

Question 3 – Strategy R&D

I then attempted to improve the strategy by tailoring the weights used for the daily portfolios. Rather than using the top and bottom decile, I instead went long the top third of stocks and shorted the bottom thirds (equally weighted). This strategy performed significantly poorer than the original:

Annualised Mean	114.43%
Annualised Standard Deviation	8.12%
Sharpe Ratio	14.09

So while the standard deviation (return volatility) did decrease, the sharp fall in returns resulted in a lower Sharpe ratio. This seems to indicate the large returns generated by the original portfolio are primarily due to investing in those stocks with very high and very low returns. The overall returns are dampened when we include those stocks with less extreme returns. This seems to contradict my previously analysis which showed that removing outliers increases the portfolio returns.

To analyse the effect of weights, and to attempt to optimise them, I would look at unequally weighting stocks in the portfolio, and analyse varying bucket proportions (top 5%, top 20%, etc).

Furthermore, noticing the downward trend in the performance of the lagged portfolios, it could be possible to increase the lag far enough such that the opposite of the signal used in this case could generate returns (a momentum strategy).