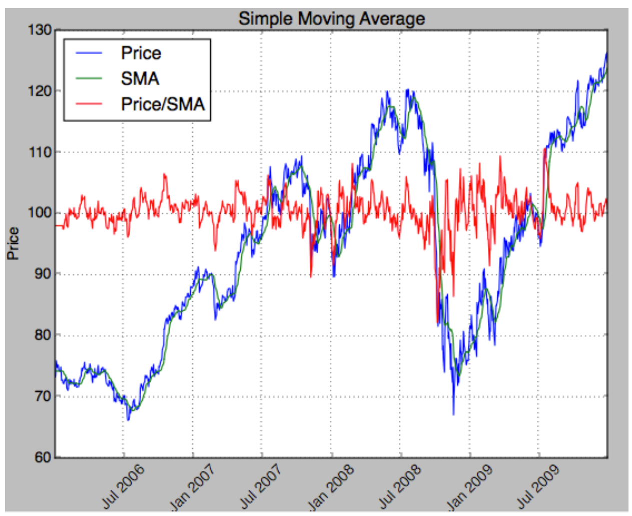
Project Report
Eliott Chapuis
GTID: 903260882

Part 1: Technical Indicators

1. Simple Moving Average (15-day)

This method takes the moving average (SMA) of stock prices over a certain lookback period (ie average over past 15 trading days) and compares it to the current price. Generally speaking, if the current price is high above the SMA, then it may fall back down soon and the opposite is true if the price is well below the SMA. To better gauge this relationship, you can use the ratio of Price/SMA, so that values above 1 indicate that the stock may be overbought and values below 1 may indicate that the stock is oversold.



15-day SMA

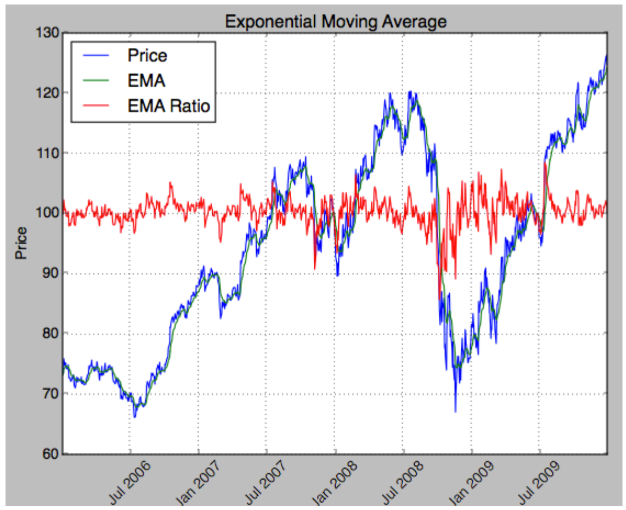
It may be hard to tell from this chart (have to keep it small), but the SMA basically follows the price of the stock with a bit of a lag time and greater smoothness. The Price/SMA ratio has larger spikes when the price changes rapidly, though there is also a lag time associated with the ratio because it is dependent upon SMA. Note that I multiplied the SMA Ratio by 100 so it can be seen more clearly.

2. Exponential Moving Average (15-day)

Similar to the simple moving average, the exponential moving average places greater weight on more recent stock prices. The weights assigned to each day are given as (1-alpha)**(n-1), (1-alpha)**(n-2)...(1-alpha), 1. Where n is the lookback period and alpha is calculated as:

$$\text{Alpha} = 2 / (n + 1)$$

The primary advantage of using EMA over SMA is that, because EMA puts greater weight on more recent stock prices, it can anticipate changes in stock trends sooner. Moreover, the ratio of Price/EMA was taken, same as with SMA, and was multiplied by 100 for scale in the following graph:



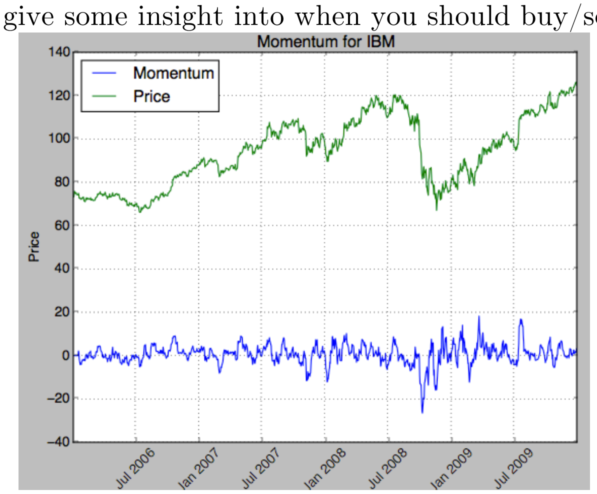
15-Day EMA

3. Momentum (10-day)

The most basic momentum indicator is simply the change in stock price over a lookback period, ie (current stock price – stock price n days ago), where n is the lookback period. However, if you want to compare the momentum value across stocks – ie between IBM and SPY – then you need to normalize somehow. You can do this by dividing the resulting momentum by past price, giving a percentage change:

Momentum Ratio = (current price – price n days ago) / price n days ago

The idea behind momentum indicators is that you want to watch for changes in momentum, ie when there is a shift from upward-trending to downward trending and vice versa. This may give some insight into when you should buy/sell.



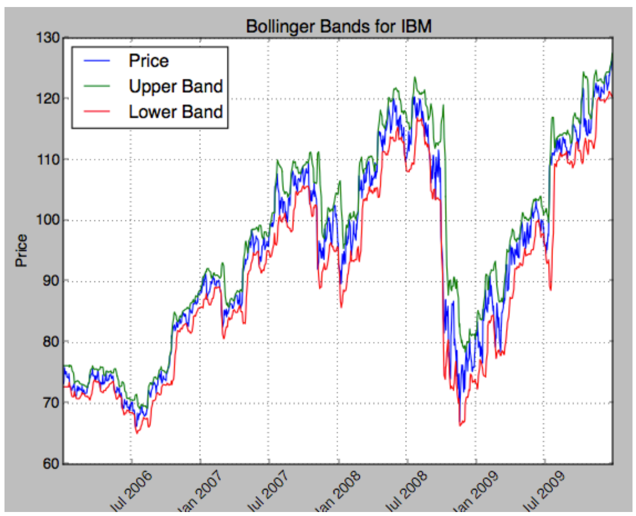
Note that momentum values were multiplied by 100 for better visualization

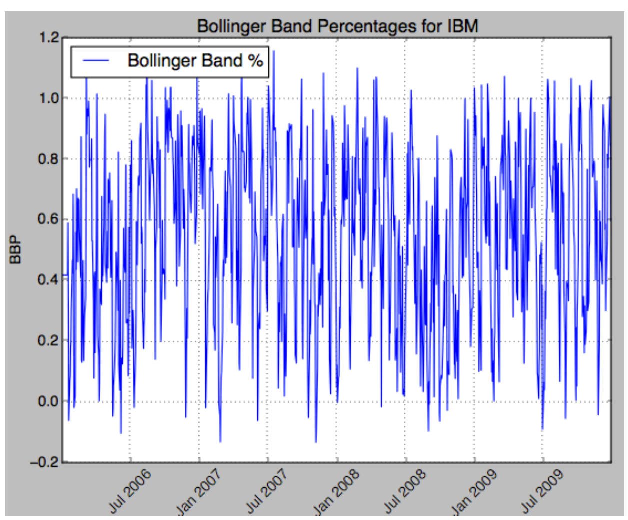
4. Bollinger Bands (15-day SMA)

This method combines a moving average with moving standard deviation to form bands around the SMA. The idea here is that if the price goes above the top band, it is probably going to fall and the opposite is true if it goes below the bottom band. To give a more sophisticated measure of the stock's relationship to the bands, you can use a percentage scale:

Bollinger Band % = (price – bottom_band) / (top_band – bottom_band)

Which gives values that are generally between 0-1 that indicate where the stock falls between the bands (lower values closer to bottom band). Values go below 0 when the stock price is below the bottom band and above 1 when the price is above the top band.





Part 2: Manual Rule-Based Trader

Constraints

We were only allowed to trade once every 10 days. To do this, I put in orders on a first come, first serve basis. So if there was another trade within 10 days of the trade before, it was simply ignored. Additionally, you could only buy or sell 500 shares every 10 days, so you can't go directly from a long to a short or vice versa. Instead, if you wanted to go from long to short, you would have to sell 500 shares one day, wait for 10 days, and sell again to get into a short position. At the very least, this makes it easier to see the black lines in the graph.

Moreover, I assessed the following rules iteratively. This means that if one of the following rules disagreed with the ones before then it would take precedent (so later rules are more important). This method did not change positions (ie change from long to short) but rather encouraged longer trading periods. Generally, I found that the SMA/EMA Ratios were the best indicators, so it makes sense that they should have priority in decision making over Bollinger Bands.

Rule 1: If the Bollinger Band % is greater than 1.1 and Momentum Ratio is greater then 0, SHORT.

So if the price goes well above the upper Bollinger Band, this is a sign that the stock may be overbought and is a strong indicator to sell. However, alone, the Bollinger Band % as an indicator does not work very well because it is a highly volatile indicator. This is where momentum comes in. This part is a bit counterintuitive, as you may want to sell when the stock is going down (momentum < 0), but the opposite is in fact better. This is because, if you believe the stock to be overbought (high BBP) then if the stock is increasing even more it will become more overbought, so you really want to short it.

Rule 2: If the Bollinger Band % is less that -0.15 and the Momentum Ratio is less than 0, GO LONG.

For this rule, I used simply the inverse logic from rule 1. In other words, you'd expect a stock price well under the lower Bollinger Band to be oversold, so you would want to go long because you expect the price to go back up. Again, momentum helps show that the stock is being even more oversold.

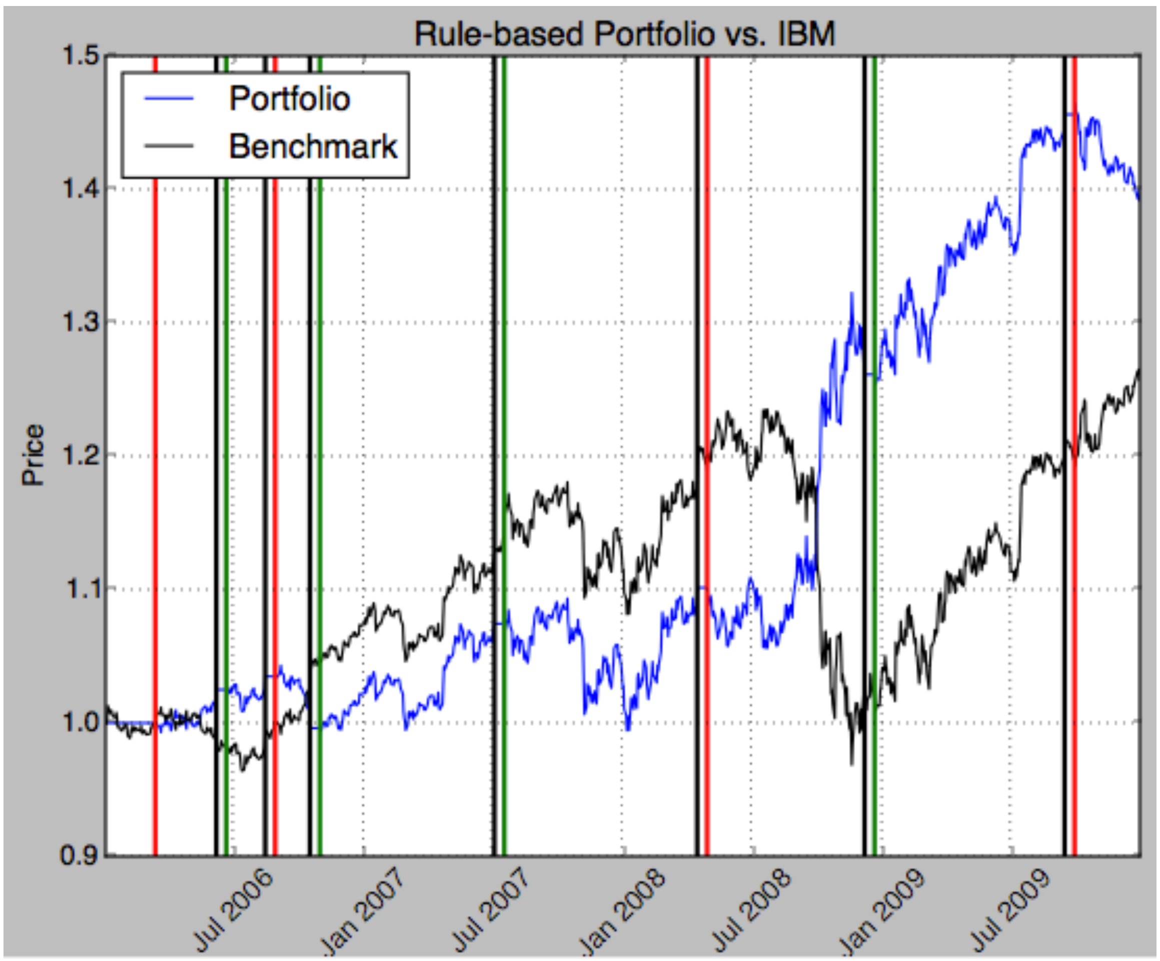
Rule 3: If the SMA Ratio (Price/SMA) is greater than 1 and the EMA Ratio (Price/EMA) is greater than 1.05, GO LONG.

Similar to Momentum, this logic is really counterintuitive, as you may expect the stock to be overpriced when the SMA or EMA ratios are higher. However, the moving averages are lagging indicators, so high ratios can actually be used to indicate a stock that is increasing in price rapidly and that has a moving average that undervalues of the true price of the stock. In other words, a high ratio may indicate that the price should be even higher than it currently is, making it a good stock to go long on. Additionally, there is a higher cutoff for EMA over SMA because EMA tends to hug the more recent stock prices more, so it is an earlier indicator of increasing stock price. In order to be more conservative about when trading decisions are made, the rule for EMA ratio is higher.

Rule 4: If the SMA Ratio is less than 1 and the EMA Ratio is less than 0.9, SHORT.

Again, this is basically the inverse of rule 3. The same indicators used to determine when a stock is going up can be used to tell if the stock is going down. In this case, you would want to short.

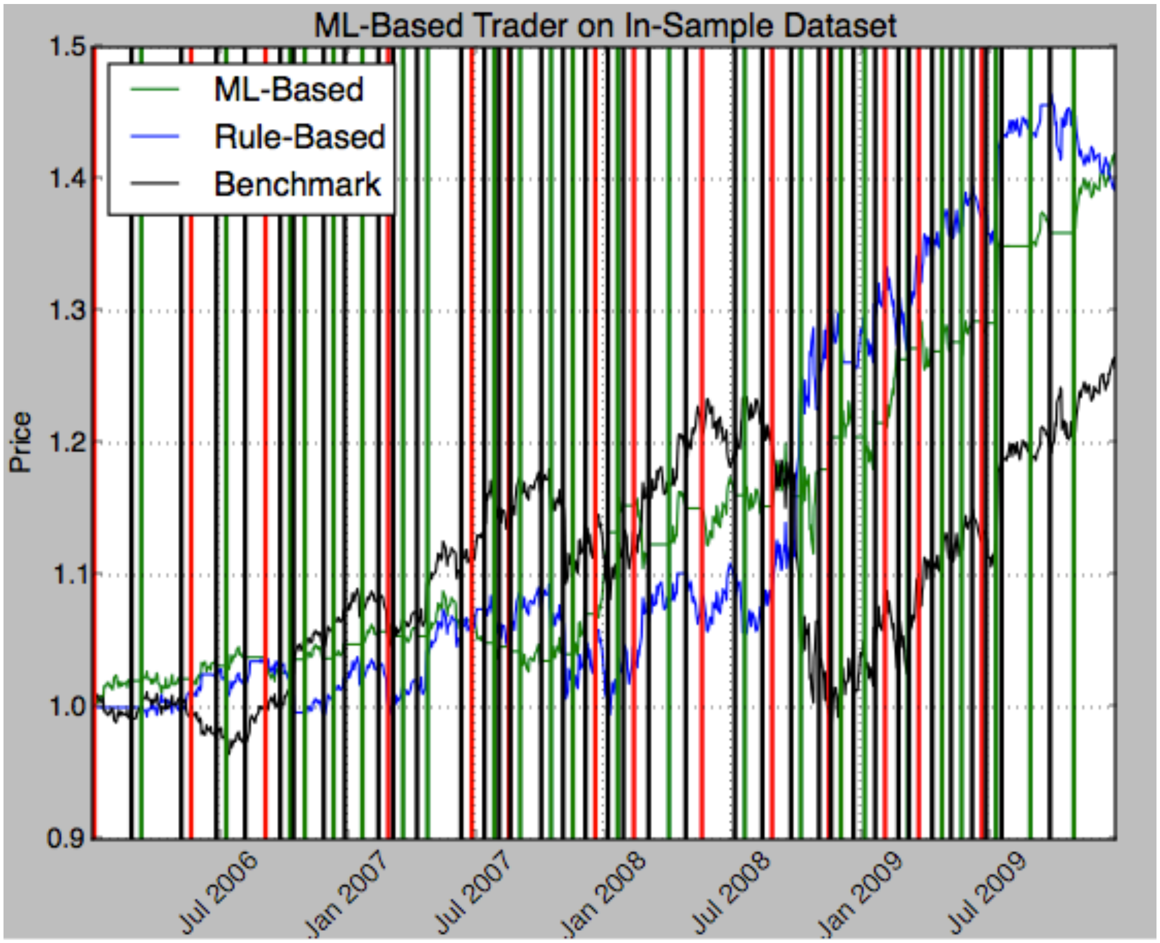
Results



**Note: the benchmark is the performance of just the individual stock, ie if you buy 500 shares at the beginning and sell at the end.

Part 3: ML Trader

Result:



ysell = -0.03, ybuy = 0.02, leaf_size = 5, bags = 30, majority = 0.4

Data

The indicators that were used included SMA Ratio, EMA Ratio, Momentum, and Bollinger Band Percentage. These values, as described in part 1, formed the X data. For the Y data, the classification value assigned to each X instance was based on the actual 10-day return. For a given day n with price p, the 10-day return would be (p_{n+10} – p_n)/p_n. If this value was higher than the 'YBUY' value (0.02) then it was classified as a +1, or a good day to go long. If the value was lower than 'YSELL' (-0.03) then the X instance was classified as -1, or a good day to go short. Otherwise, do nothing.

Learner

The learner used was a simple Random Tree Learner, with bagging.

Parameters

So all of the variables are as defined in the wiki, with the exception of majority. When bagging, the average of the learner decisions is calculated. So for example, if the average is above 0.33, then a majority of learners said go long. If the value is below -0.33, then a majority of learners said go short. By putting the majority value at 0.4, you raise the bar a little bit by saying that a super-majority of sorts is needed. I tested for a bunch of values and 0.4 seemed to work decently. In this way, you can translate the bagging learner results (ie the average of the RTLearners) into meaningful classifications based on the number of RTLearners selecting each option. Additionally, to achieve these results the parameters used were a leaf size of 5 and 30 bags.

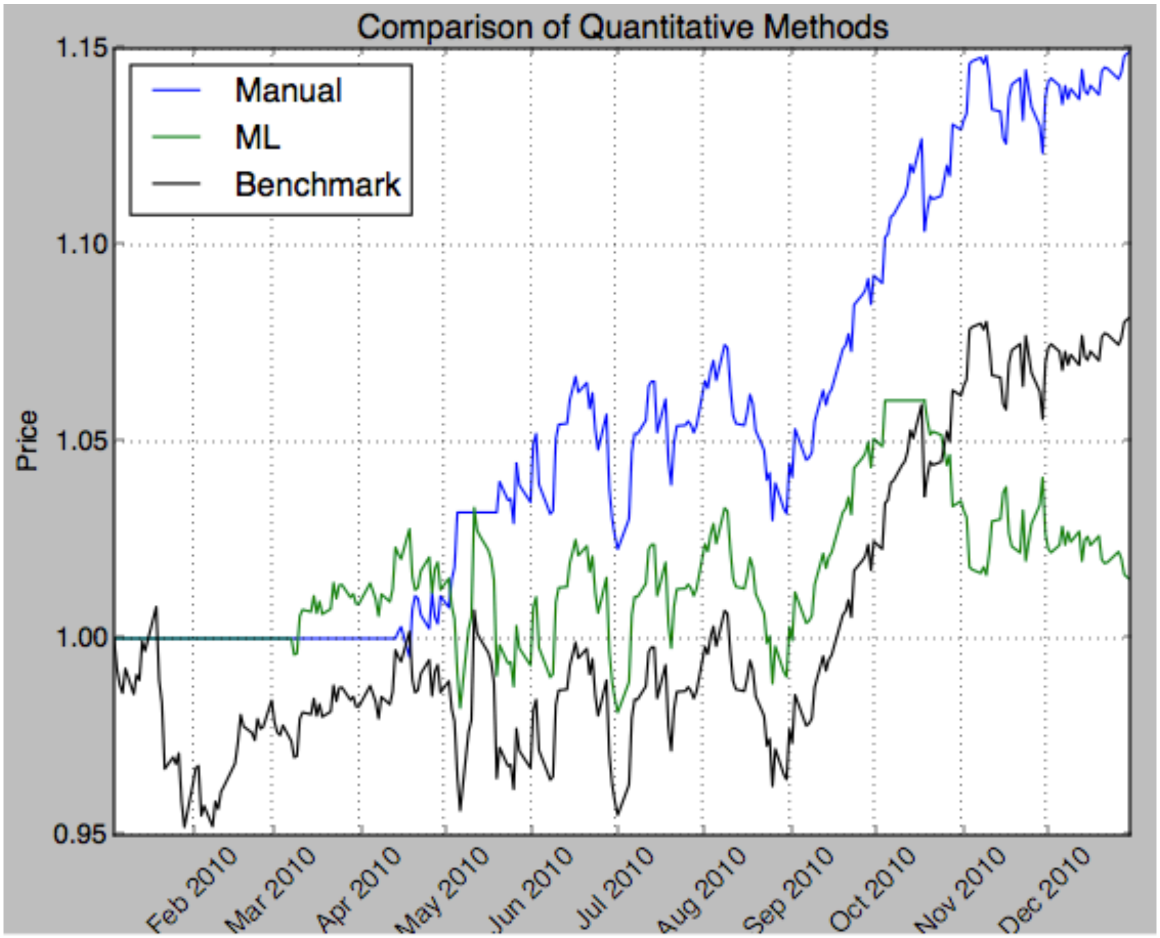
Additional Comments

It's very important to note that these results are not very consistent. Because the ML strategy is dependent upon RT Learners, the results may vary. While the cumulative return is often better than the benchmark, it is not always above 1.5 times benchmark. However, this example demonstrates that the model is in fact capable of achieving good results.

Additionally, I'd like to point out that the ML strategy does actually give a slightly higher cumulative return than the rule-based strategy.

Part 4: Comparative Analysis

Results:



Parameters for ML Strategy: ysell = -0.03, ybuy = 0.02, leaf_size = 5, bags = 30, majority = 0.4 (Same as before)

	Training Dataset			Testing Dataset		
	Manual	ML	Benchmark	Manual	ML	Benchmark
Sharpe Ratio	-0.7998	-1.1534	-0.4774	-1.6353	-0.1066	-0.6493
Cumulative Return	39.99%	41.13%	25.72%	14.92%	1.52%	8.16%
Standard Deviation	0.0062	0.0046	0.0068	0.0052	0.0061	0.0070
Average Daily Return	-0.031%	-0.033%	-0.020%	-0.054%	-0.004%	-0.029%
Final Portfolio Value	139985	141125	125715	114915	101515	108155
Correlation (with Benchmark)	0.42519	0.56995	N/A	0.90618	0.67847	N/A
RMSE (with Benchmark)	0.03353	0.03231	N/A	0.05595	0.03092	N/A

Discussion:
So the first thing that you notice in the graph is that the performance of the ML strategy is worse than both the benchmark and the Manual Strategy. This largely has to do with the decisions made for training on the in-sample data. The small leaf size (5)

means that overfitting of the training data is probably going to occur, decreasing the generalizability of the ML Strategy on test data.
It is probably for this reason that the manual strategy works better over test data because the rules applied to the in-sample data can also apply to the out of sample data. However, it is important to mention that the Manual Strategy only makes a couple of trades over the testing period, which is why it looks a lot like the benchmark line.

Metrics
In some ways, the ML strategy may be preferable because it is less correlated with the Manual strategy. The Manual strategy really takes advantage of moving averages, so it tends to buy in when the stock is moving up and short when it is going down, without really predicting the change, just mirroring it with some lag time. However, ultimately the feature that I looked at the most when evaluating performance was the final portfolio value or cumulative return because that's really what counts. Looking at this, the manual strategy obviously performed much better and I believe that this is due to the generalizability of the rules. That being said, it would be interesting to see what would happen with both strategies over a longer period of time.