Project 6: Indicators and TOS Report

William Callan
wcallan3@gatech.edu

***Abstract*—A manual training strategy was compared to a Q-Learner strategy for maximizing profit during a trading period.**

1 INDICATOR OVERVIEW

1.1 SMA

1.1.1 *Implementation*

Simple Moving Average (SMA) for a stock is calculated by choosing a window size and taking the mean of the previous *window* days of the stock's price. To reduce the data down to a single vector and compare it to the stock price, at each data point the value (*price of stock* – SMA) was found. Then, to standardize these values, these data were divided by the standard deviation of the entire SMA period: the formula is $\frac{prices-SMA}{std(SMA)}$.

1.1.2 *Parameters*

Window size of the SMA needed to be optimized. A value of 20 was chosen as it offered acceptable smoothing of the noisy data while also not pulling from data that was more than a month old for a stock. MORE?

1.2 Momentum

1.2.1 *Implementation*

Momentum is calculated by taking a day's price and dividing it by the price from *window* days ago, then subtracting 1: the formula is $\left(\frac{current\ price}{older\ price}\right) - 1$. Then, to standardize these values, these data were divided by the standard deviation of the entire Momentum vector.

1.2.2 *Parameters*

Window size for determining the older stock price needed to be optimized. A value of 20 was chosen because it matched the value chosen for SMA and the

manual strategy relied on multiple indicators giving a signal for the same data point.

1.3 PPO

1.3.1 *Implementation*

There are two variables calculated with PPO: the PPO and the signal line.

Percentage price oscillators (PPO) leverage the exponential moving average (EMA) of a stock. To calculate EMA, first calculate a multiplier as $\frac{2}{window\ size+1}$. The formula for EMA of the current day is $current\ price * multiplier +$ EMA(yesterday) $* (1 - multiplier)$.

Choose two window sizes, a smaller and a larger value. Then calculate the EMA for the stock using both of these window sizes, resulting in EMA_{small} and EMA_{large}. The formula for the PPO line is $\frac{EMA_{small} - EMA_{large}}{EMA_{large}} * 100$. To calculate the signal line, take the EMA of the PPO line with a third window size. To reduce this to a single standardized vector, take the difference between the PPO line and the signal line and divide it by the standard deviation of this difference: the formula is $\frac{PPO-signal}{std(PPO-signal)}$.

1.3.2 *Parameters*

Three window sizes needed to be optimized: the small window for the PPO, the large window for the PPO, and the window for the signal line. For the PPO, values of 12 and 26 were chosen: this was to most closely match the value of 20 chosen for the SMA and Momentum indicators with 12 and 26 being on either side of this value. The value of 9 was chosen for the signal line window: this value offered acceptable smoothing of the noisy data while also not pulling from data that was more than two weeks old for the PPO.

2 MANUAL STRATEGY

2.1 Trade signals

The SMA, momentum, and PPO vectors calculated in section 2 were used to determine trade signals. When a vector crossed from negative to positive, a value of 1 was ascribed to that vector; when a vector crossed from positive to negative,

a value of -1 was ascribed; if no sign change occurred, a value of 0 was ascribed. These three values were added together: if the sum was positive (i.e., at least one indicator changed to a positive sign) a buy signal was issued, if the sum was negative a sell signal was issued, and if the sum was 0 then a hold signal was issued.

The sign change of these vectors was chosen because each of these indicators is a measurement of the general trend of a stock, and when a sign change occurs it is an indication of a potential uptrend or downtrend in the stock's immediate future. The hypothesis is that when multiple indicators have complementary sign values (e.g., multiple indicators changed from positive to negative on the same day) the likelihood of a trend emerging is heightened, hence a trade signal is issued.

2.2 Performance

For the in-sample period, the benchmark trading strategy for JPM stock yielded a cumulative return of 0.012325, while the manual trading strategy yielded a cumulative return of 0.411765; this can be seen in Figure 1.

For the out-of-sample period, the benchmark trading strategy for JPM stock yielded a cumulative return of -0.083579, while the manual trading strategy yielded a cumulative return of -0.205234; this can be seen in Figure 2.

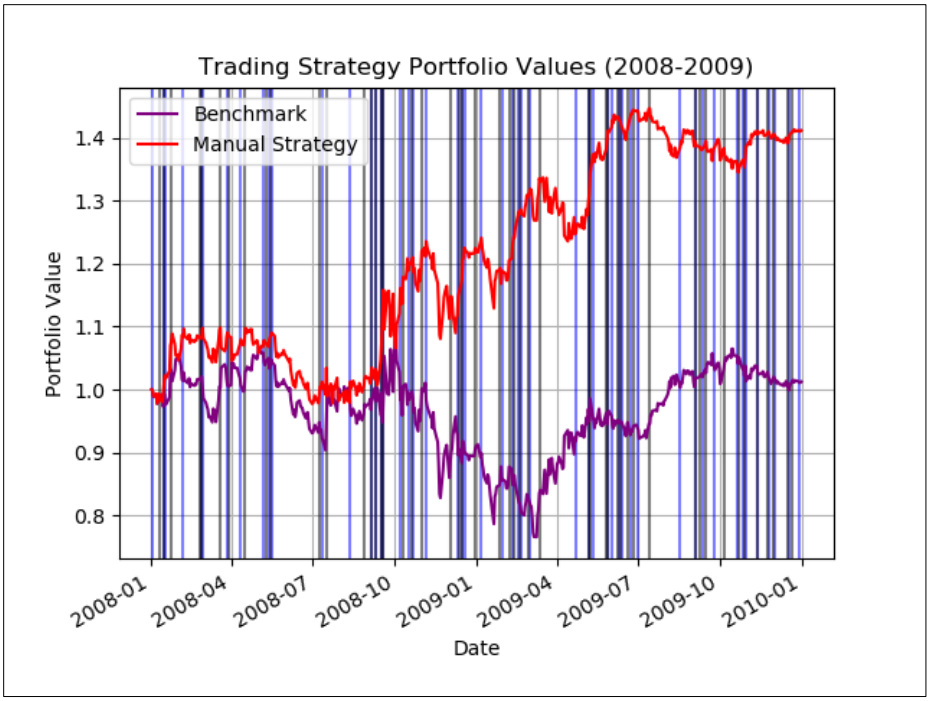

Figure 1 — Comparing portfolio values of benchmark trading strategy to manual trading strategy of in-sample period.

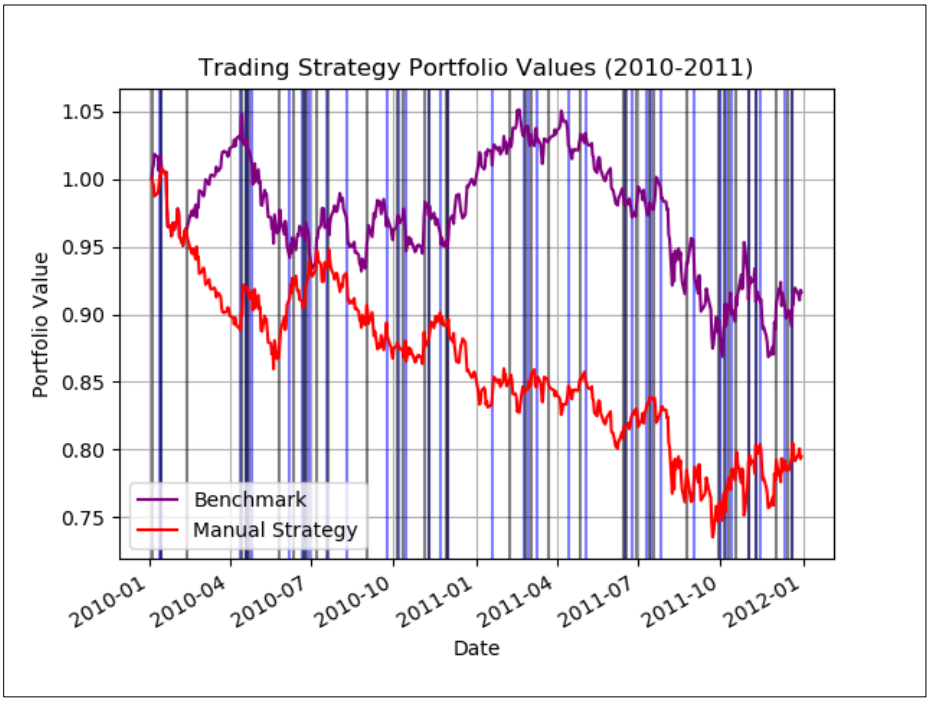

Figure 2 — Comparing portfolio values of benchmark trading strategy to manual trading strategy of out-of-sample period.

2.3 Evaluation

For the in-sample period, the manual strategy significantly outperformed the benchmark. However, it performed worse than the benchmark in the out-of-sample period. This could be an indication that the out-of-sample period experienced more volatile stock movement than the in-sample. The manual strategy relies on continuous trends in a stock's momentum, so a rapidly fluctuating stock period would yield poor results with this trade strategy.

3 STRATEGY LEARNER

A Q-learner was used as the strategy learner. It was trained on the same three indicators as the manual strategy: single vector versions of SMA, momentum, and PPO.

3.1 Framing the Problem

The state of the stock market was discretized with 4 digits:

- Whether the agent had holdings of 0, 1000, or -1000 shares. These were represented as the digits 0, 1, and 2.
- Current day's (price-SMA). The whole range of values was discretized and represented by the digits 0, 1, 2, 3, 4.
- Current day's momentum. The whole range of values was discretized and represented by the digits 0, 1, 2, 3, 4.
- Current day's (PPO-signal). The whole range of values was discretized and represented by the digits 0, 1, 2, 3, 4.

These digits were then combined into a single integer, resulting in 375 distinct states ($3 * 5 * 5 * 5$).

The learner could take one of three actions at each state: hold, long until holdings is 1000 shares, and short until holdings is -1000 shares. This resulted in a Q-table with dimensions [375, 3].

The reward function for the learner was the daily return after taking an action: this can be calculated by taking the portfolio value of the day after taking the action, dividing it by the portfolio value of the day prior, then subtracting 1: the formula is $\frac{value_N}{value_{N-1}} - 1$.

3.2 Hyperparameters

For the below hyperparameters, the number in parentheses indicate the value used for the strategy learner.

- Indicator discretization (5): For determining the state, each indicator was represented by a single digit having one of N values. 5 was chosen for N because with any more digits the number of states would have exceeded the number of data points present in the training set, guaranteeing some states could never be visited in the training.
- Learning Rate (0.2): How quickly the Q-table is updated when presented with new reward values. 0.2 was chosen because other values yielded a lower in-sample performance result.
- Discount Rate (0.9): How much impact future rewards have on an action's current reward.0.9 was chosen to mimic real-world evaluations of future reward for stocks.
- Random Action Rate (0.5) and Random Action Decay Rate (0.99): How often the Q-learner chooses a random action for a state and how quickly the random action rate decays to 0. 0.5 and 0.99 were chosen to maintain random actions for the first iteration only; these values resulted in a random action rate of around than 0.004 by the end of the first iteration of the Q-learner, resulting in future iterations having essentially 0 random actions taken and instead focusing on convergence of the model.
- Dyna (300): The number of hallucinated experiences the learner creates after each observation. 300 was chosen to keep the training time under 25 seconds and because higher values converged to the same table and as such were unnecessary.
- Iterations (3): The number of times the Q-learner was updated on the training set. 3 was chosen to keep the training time under 25 seconds, and since dyna was used a larger number of iterations rarely affected the model's convergence.

3.3 Data standardization

The process for standardizing the three indicators is described in sections 1.1.1, 1.2.1, and 1.3.1.

4 EXPERIMENT 1

4.1 Methods

The cumulative returns of the manual strategy and strategy learner were compared to a benchmark trading strategy of longing 1000 shares on day 1 and holding for the entire period. The exact parameter values used for the strategy learner are described in section 3.2. The strategy learner was trained on the in-sample period of 2008-1-1 to 2009-12-31 for the stock JPM. After training, the strategy learner was queried on both the in-sample and out-of-sample sets.

The hypothesis is that the strategy learner should produce a higher cumulative return than both the benchmark and manual strategies for both the in-sample and out-of-sample periods. This is because the strategy learner should be able to find more complex patterns within the indicators' data than the manual strategy could.

4.2 Results

Figure 3 shows the performance of the strategies for the in-sample dataset and Figure 4 shows the performance for the out-of-sample dataset.

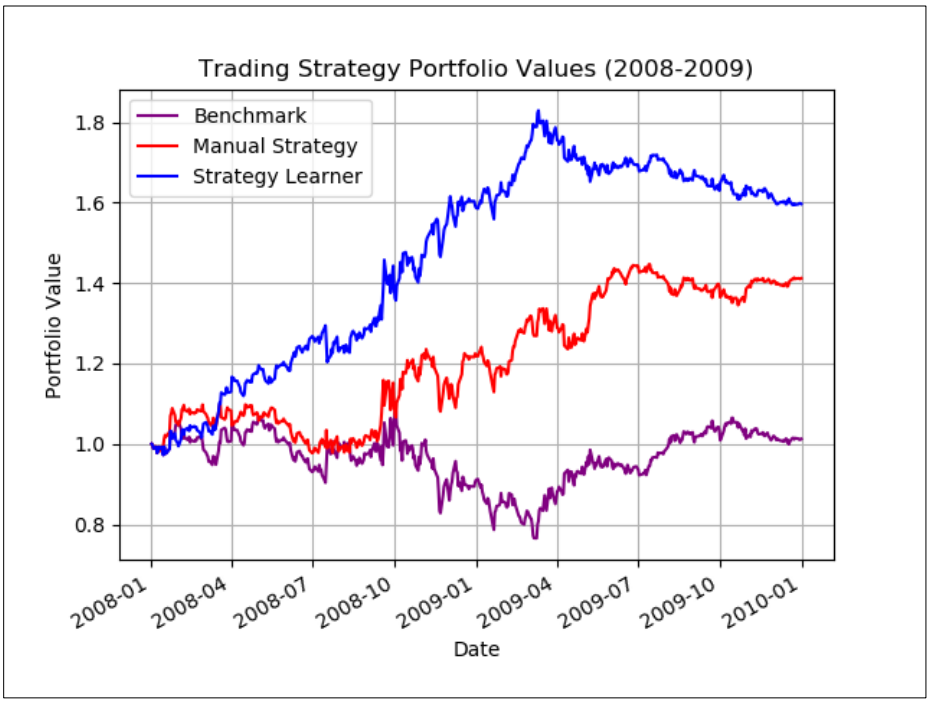

Figure 3 — Trading strategy comparison for in-sample period.

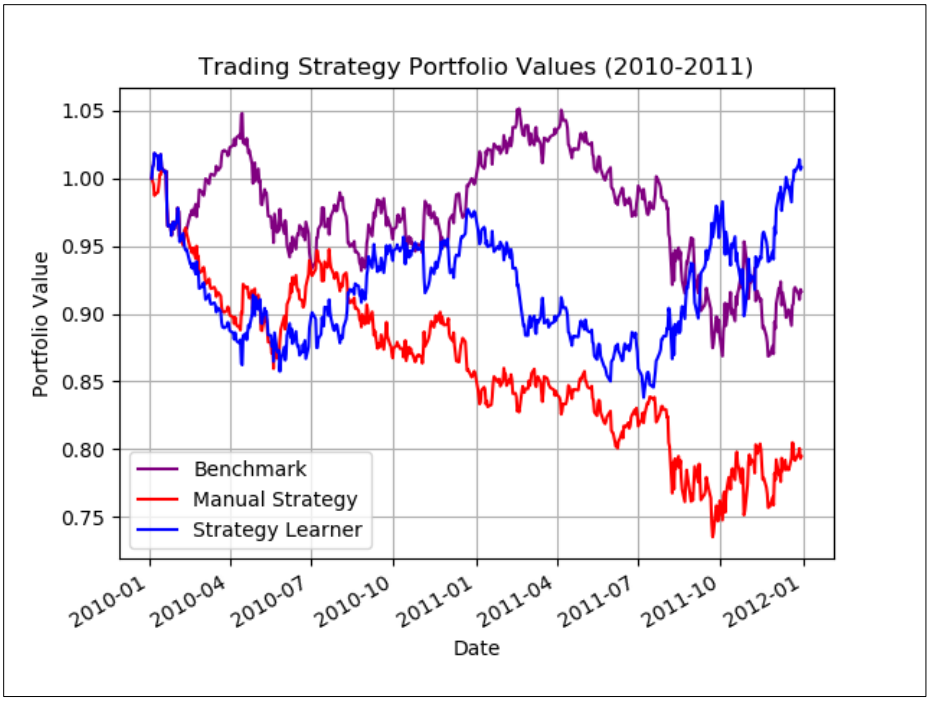

Figure 4 — Trading strategy comparison for out-of-sample period.

As seen in the figures, the hypothesis was proven correct with the strategy learner achieving cumulative returns of 0.596763 for the in-sample set and 0.008005 for the out-of-sample set, higher than both the benchmark and the manual strategy. This is to be expected for the in-sample data, as the strategy learner is allowed to repeatedly attempt to maximize its rewards on this dataset, therefore fitting the data quite well.

5 EXPERIMENT 2

5.1 Hypothesis

The hypothesis is that the strategy learner will achieve a lower portfolio value as the impact increases and also that the strategy learner will make fewer trades as the impact increases. This is due to each trade reducing the portfolio value by a higher amount as impact increases.

5.2 Methods

Three strategy learners were created with the same parameter values described in section 3.2 and trained on the same dataset: JPM from 2008-1-1 to 2009-12-31. However, each learner was initialized with a different impact cost: one had an

impact of 0, the second had an impact of 0.0075, and the third had an impact of 0.0150.

5.3 Results

Figure 5 shows the portfolio value after applying the strategy learner. The hypothesis was supported as impact increasing caused a clear drop in final portfolio value.

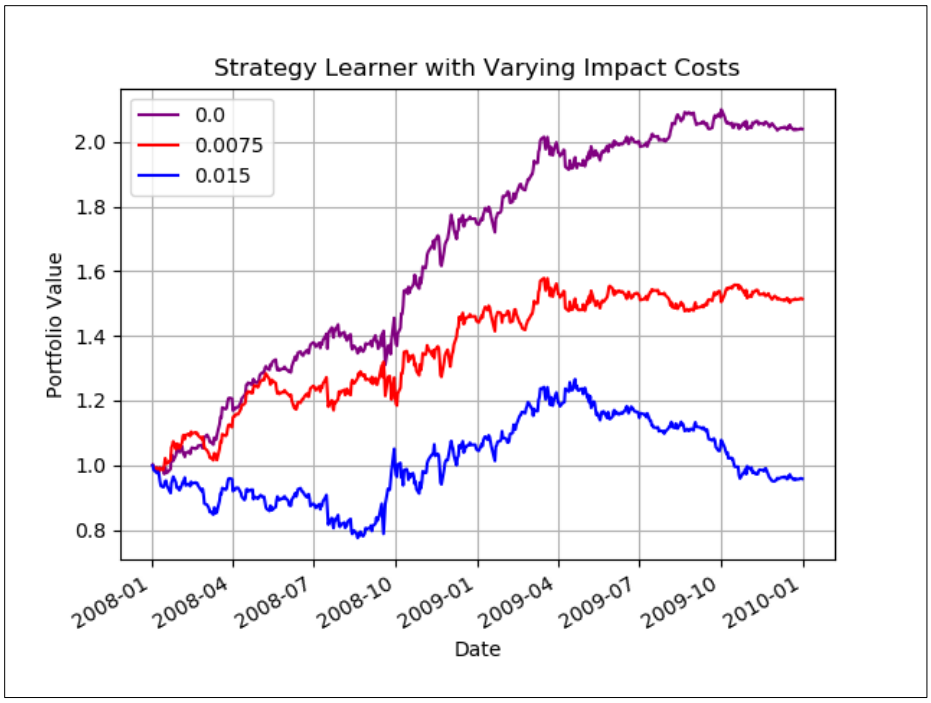

Figure 5 — Portfolio values of strategy learner being trained on the same dataset with varying impact values.

Table 1 details how many trades were initiated by each strategy learner. Surprisingly, the impact value seemed to have no correlation to the number of trades. This may be because the indicators that the learners trained on weren't affected by impact at all and thus the learner would have trouble determining how its actions affected the state. The only influence that impact had over the learner was that it affected the magnitude of the reward function values, and without many more iterations these lower rewards would remain uncaught by the learner. More testing with a higher iteration count would need to be conducted to possibly reveal a correlation between number of trades and impact value.

Table 1 — Varying impact values and the number of trades that the strategy learner issued.

Impact: 0.0000	**Impact: 0.0075**	**Impact: 0.0150**
72	47	72