Strategy Evaluation

Chen Jiang
cjiang98@gatech.edu

1. INTRODUCTION

In this project, I used historical stock prices to optimize the timing of buying and selling, and evaluated the good and bad of both methods. And evaluated the advantages and disadvantages of Manual Strategy and Strategy Learner.

2. INDICATOR OVERVIEW

2.1 Simple Moving Average (SMA)

A simple moving average (SMA), also known as an "arithmetic moving average", is a simple averaging of closing prices over a specific period. The term "moving average" refers to a simple moving average (SMA).

$$SMA = (C1+C2.....+Cn)/n$$

N means the look back period (the number of the rolling days). A crossover of SMA and the price can be used to signal a change in trend and can be used to trigger a trade.

2.2 **BBP%**

The Bollinger Bands indicator is derived from statistical averages and incorporates two standard deviations above and below the mean. According to the Bollinger Bands theory, the price of the underlying asset tends to remain within the Bollinger Bands. When it strays from this range, it is expected to eventually retrace back to the Bollinger Bands over the long term.

Mid line = N-Day Price Average

*Upper line = Mid line + 2 * N day price standard deviation*

$$Lower\ line = Mid\ line - 2 * N\ day\ price\ standard\ deviation$$

Bollinger Bands Percentage(BBP%) combines the upper and lower band,

$$BBP\% = \frac{Price - lower\ Line}{upper\ line - \ lower\ line}$$

2.3 WILLIAMS

The Williams %R indicator is a technical indicator used to measure overbought and oversold conditions in stocks or other financial assets. It was developed by Larry Williams to help identify price reversal points and trading signals.

$$Williams\ \%R = (Hn - C) / (Hn - Ln) * -100$$

3. MANUAL STRATEGY

3.1 Describe how you combined your indicators to create an overall signal.

STEP1: In this section, I used SMA, BBP and Williams %R to combine to determine the time of the trade. SMA is a ratio that indicates the price's relative position. A low ratio indicates that it is possible to buy. Conversely, a high ratio indicates a time to sell. BBP (Bollinger Band Percentage) is a technical indicator. Generally above 0.8 investors can sell and below 0.2 can buy. Williams %R. When the indicator exceeds -20, it suggests that the asset is overbought, indicating a potential price reversal or correction. Conversely, when the indicator falls below -80, it suggests that the asset is oversold, indicating a potential price reversal or rally.

In this assignment, I used a combination of these indicators. By obtaining different indicator thresholds to determine the timing of buying or selling. There is no good way to determine the optimal threshold. Only manual and slow adjustment.

STEP2: **Adjustment parameters**. In practice, the value of these indicators is difficult to judge, I kept adjusting based on the in-sample data. The following are some of the results of my adjustments. (It may be possible to find some way to automatically tune the parameters, but I did not do so in my experiments this time.)

combinations	Williams= -95/-5 SMA=0.4/1.1 BBP=0.2/0.8	Williams=-80/-20 SMA=0.3/1 BBP = 0.25/0.75	Williams=-90/-10 SMA=0.4/1.1 BBP= 0.3/0.7	Williams=-85/-15 SMA=0.4/1.1 BBP = 0.25/0.75
Sharpe Ratio	1.264682	0.344321	0.798293	0.069827
Cumulative Return	0.55702	0.089734	0.338728	0.239829
Mean of Daily Return	0.000949	0.000302	0.000592	0.000590
Final Portfolio Value	155686.6	108930.1625	128713.25	118945.75

Table 1— Manual Strategy parameters performance (In sample)

From the above experiment, it can be seen that Williams=-95/-5, SMA=0.4/1.1, BBP =0.2/0.8 are the best combination of indicators.

STEP3: A sell signal is generated when the price data reaches the upper line of the above indicator. A buy signal is generated when the price data reaches the lower line of the above indicator.

STEP4: Create follow-up operations based on trading signals and current positions.

In this experiment, the trading rules are set as follows:

In sample：Symbol: JPM，Date：January 1, 2008 to December 31, 2009, starting cash: $1000000, Commission: 9.95, Impact: 0.005

Out of sample：Symbol: JPM，Date：January 1, 2010 to December 31 2011, starting cash: $1000000, Commission: 9.95, Impact: 0.005

Others: 1000 shares long, 1000 shares short, 0 shares. All results are accurate up to six decimal places.

3.2 Compare the performance of your Manual Strategy versus the benchmark for the in-sample and out-of-sample time periods.

In Sample Results:

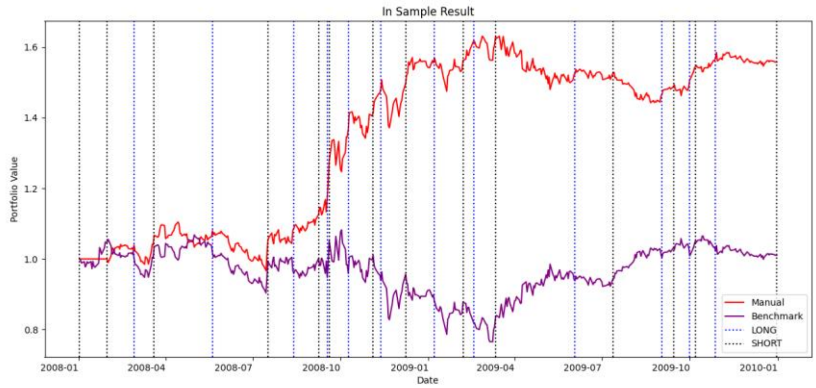

Figure 1— in sample results

	Manual	Benchmark
Daily Return	0.000949	0.000164
Cumulative Return	0.55702	0.010236
Sharpe Ratio	1.264682	0.153386
STDEV	0.011918	0.017041

Table 2— in sample results

Out of Sample Results:

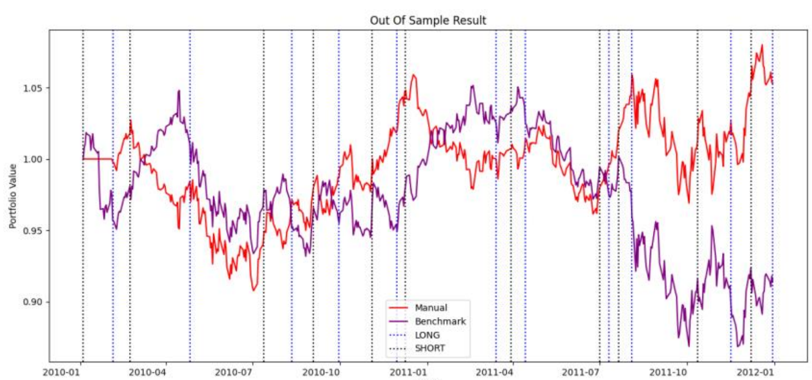

Figure 2— Out of sample results

	Manual	Benchmark
Daily Return	0.000133	-0.000141
Cumulative Return	0.053289	-0.085309
Sharpe Ratio	0.273054	-0.263617
STDEV	0.007737	0.008501

Table 3—Out of sample results

3.3 Evaluate the performance of your strategy in the out-of-sample period.

Referring to Figure-1, it is evident that the Manual strategy exhibits superior performance compared to the benchmark throughout the sample period. The cumulative return of the Manual strategy significantly surpasses that of the benchmark. Additionally, the Manual strategy demonstrates better results in terms of the standard deviation and mean of daily returns. These findings indicate that the Manual strategy outperforms the benchmark across multiple measures.

Analyzing Figure-2, it becomes apparent that the returns, while not as favorable as the In Sample results, still showcase the outperformance of the manual strategy compared to the benchmark in the Out of Sample period. The cumulative return of the manual strategy exceeds that of the benchmark. Furthermore, other metrics also exhibit better performance for the manual strategy. But the performance was much less than the in-sample tests.

3.4 WHY these differences occur.

The discrepancy in performance between the in-sample and out-of-sample periods can be attributed to a straightforward reason: our manual model was solely optimized using in-sample data, without considering out-of-sample data. As in Table1, all parameters were repeatedly adjusted. As a result, the model generated from the in-sample data may not be well-suited for the out-of-sample data, leading to its underperformance. This outcome is understandable given the mismatch between the model and the out-of-sample data.

4. STRATEGY LEARNER

4.1 Describe the steps you took to frame the trading problem as a learning problem for your learner.

STEP1: In this section, I used a classification-based learning approach to handle buy and sell decisions, specifically using RTLearner and BagLearner.

STEP2: I utilize the same indicators as those used in the manual strategy, namely SMA, BBP, and Williams %R. I made sure that the parameters were consistent with previous experiments. These indicators were combined in a table after they were calculated from prices.

STEP3: After that, I use the SMA, BBP, and Williams %R values computed within the in-sample dataset, along with the set of indicators generated based on price changes, as training INPUTS FEATURES(X) for the learner. Trading operations are generated according to the set of indicators and consist of three categories: Long (+1), Short (-1), and Cash (0). These operations are the OUTPUT TARGET(Y). In order to avoid the increased costs of excessively frequent transactions, a trigger condition is established, the following formula used.

$$BUY = 0.02 + impact \,/\, SELL = -0.02 - impact$$

Step 4: Based on the above data, we can input X and Y to generate the model. After the training is completed, the new stock dataset is imported into the model to get the corresponding set of trades.

4.2 Describe the hyperparameters, their values, and how they were determined.

In this experiment, I used random forests as a model for machine learning. Leaf Size=5 and bags=20. The indicators I use are SMA, BBP, and Williams %R. In the training, I use the 5-day returns of the stocks during the sample period to determine the strategy. The leaf size and the number of the bag are very important parameters. In this experiment, I experimented with different parameters to determine which combination would give the most optimal results. In my experiments, Leaf Size=5 and bags=20 gave the best results.

I am not using Q-Learn, so there is no need for discretization.

5. EXPERIMENT 1

5.1 Describe your experiment in detail

Assumption: Using machine learning will result in better performance.

I conducted a comparison between the MANUAL STRATEGY, STRATEGY LEARNER, and the BENCHMARK. The objective was to assess the performance of these strategies by examining the results through three metrics: Daily Return, Cumulative Return, and STDEV. By analyzing the outcomes of the three strategies, I aimed to evaluate their respective effectiveness.

The stock used is JPM, the time period is January 1, 2008 to December 31, 2009, the starting cash is $100000, the Commission is 9.95 and the Impact is 0.005. In the benchmark, buy 1000 shares at the beginning and sell 1000 shares at the end.

	Strategy	**Manual**	**Benchmark**
Daily Return	0.001405	0.000949	0.000164
Cumulative Return	0.952765	0.55702	0.010236
Sharpe Ratio	1.807431	1.264682	0.153386
STDEV	0.012335	0.011918	0.017041

Table 4—Experiment 1 result (In sample)

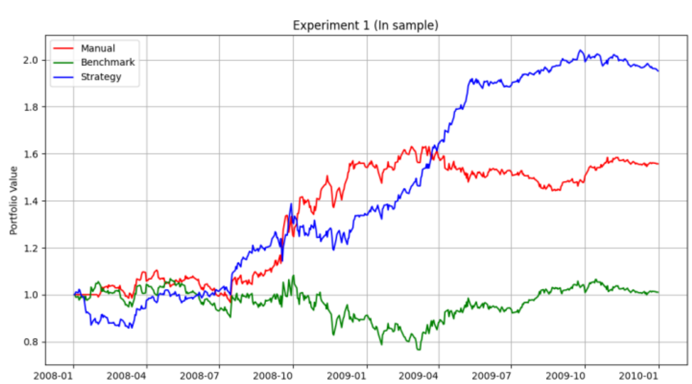

Figure 3—Experiment 1 result

5.2 Describe, interpret, and summarize the outcome of your experiment

Based on the provided graphs, a clear distinction can be observed among the strategy learner, manual strategy, and benchmark. Within the sample, the strategy learner demonstrates the highest return, while the manual strategy achieves a respectable return due to certain manual adjustments. Conversely, the benchmark exhibits the lowest return. These observations highlight the considerable differences between the three approaches.

From the results of my experiments, in-sample strategies do perform better, reflecting the power of machine learning. But I don't think this will be the case every time. Because of the presence of overfitting, Data Bias, algorithm selection and hyperparameter tuning. All these factors may affect the performance.

5.3 Out-of-sample experiment

	Strategy	**Manual**	**Benchmark**
Daily Return	-0.001108	0.000133	-0.000141
Cumulative Return	-0.443085	0.053289	-0.085308
Sharpe Ratio	-1.683018	0.273054	-0.263617
STDEV	0.010454	0.007737	0.008501

Table 5—Experiment 1 result (In sample)

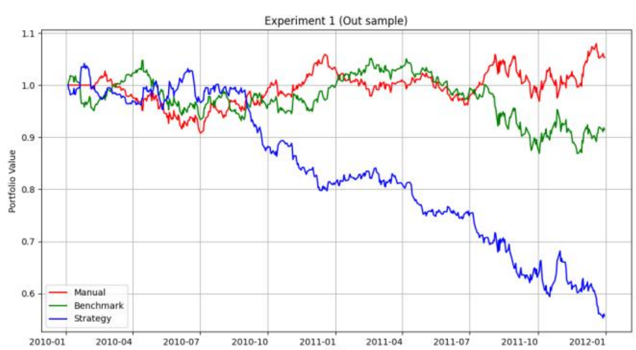

Figure 4—Experiment 1 result (Out of Sample)

As can be seen from the results of the experiment, in the out-of-sample test, the manual strategy performs the best, the benchmark follows, and the strategy learning is the worst. This result shows that great care still needs to be taken when using machine learning. **Machine learning is not a silver bullet.**

This result is the exact opposite of the initial hypothesis. There should be many reasons for this, such as the choice of indicator group, the choice of model, and the setting of hyperparameters. In short, the out-of-sample performance of my model is poor. If it is put into use, I think I will go bankrupt.

6. EXPERIMENT 2

Assumption: High impact will reduce the return.

In experiment 2, I employed the strategy learner to investigate the impact of different values of IMPACT (**0**, **0.005**, and **0.015**) on trading outcomes. The metrics used for evaluation were Daily Return, Cumulative Return, and Sharpe Ratio. The experiment was conducted over the period from January 1, 2008, to December 31, 2009, starting with a cash balance of $100,000.

For this particular section, the parameters utilized were a Leaf size of 5, a Bag Size of 20, and a moving window of 21 days. The combination of learned indicators consisted of SMA, BBP, and Williams %R. The benchmark portfolio began with $100,000 in cash and invested in 1,000 shares of the specified symbol on the first trading day. These shares were subsequently sold on the last trading day.

The following are the results of Experiment 2

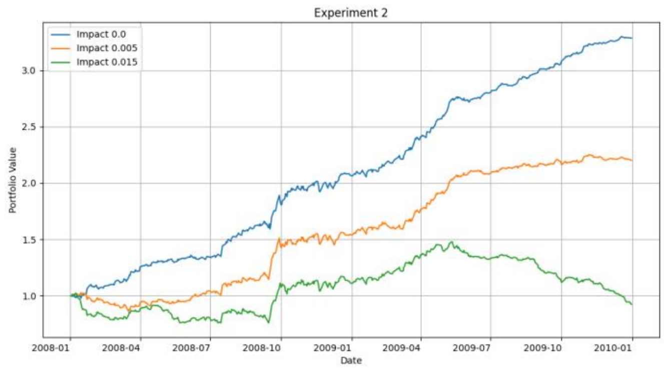

Figure 5—Experiment 2 result with different impact

Impact	**0**	**0.005**	**0.015**
Daily Return	0.002402	0.001632	-0.000045
Cumulative Return	2.2849	1.200506	-0.07675
Sharpe Ratio	4.26354	2.255558	-0.047744
STDEV	0.008944	0.011483	0.015089

Table 6—Experiment 2 result with different impact

Based on the provided chart, there is a clear negative correlation between Impact and Final return. As Impact increases, Sharpe Ratio, Daily return, and Cumulative return all decrease. This can be attributed to the fact that Impact represents transaction costs, and higher costs are expected to result in lower returns.

Impact	**0**	**0.005**	**0.15**	**0.3**	**0.4**
Number of trades	146	130	83	24	1

Table 7—Experiment 2 Number of trades with different impact

Additionally, an increase in Impact leads to a decrease in the volume of trades. This negative relationship between price impact and trading volume suggests that as the impact increases, the volume of transactions declines.