Machine learning for Trading (CS7646)

Indicator Evaluation

Shahrzad Amoozegar
shahrzad.amoozegar@gatech.edu

This project focuses on developing a strategy for trading stocks. Two different strategies including a manual strategy and a classification-based learner have been introduced and compared. These strategies utilize technical indicators to make trading decisions.

Indicators:

I used 3 different indicators in this project for both manual strategy and classification-based learner. These indicators are briefly explained below:

1.Momentum: Momentum is the rate of price change. It shows the movement of price and strength of the movement over time.

Rate of change $= \frac{\boldsymbol{price_{today} - price_{N\ days\ ago}}}{\boldsymbol{price_{N\ days\ ago}}} = \frac{price_{today}}{price_{N\ days\ ago}} - 1$

2. Price/SMA (price to simple moving average ratio): Price to SMA ratio shows the price of the stock at day t divided by simple moving average over the last N periods

price/SMA formula:

SMA (simple moving average): SMA is the sum of prices over a number of periods divided by the number of time periods.

SMA $= \frac{\sum_N price}{N}$ Price/SMA $= \frac{price_t}{SMA_{N\ periods}}$

3. Bollinger bands Percentage: A Bollinger band is an indicative that provides a range within which the stock price trades. Bollinger bands consist of a centerline (exponential moving average) and two price bands (standard deviation of the stock prices) above and below it.

Upper Bollinger band: SMA + 2*standard deviation of stock prices

Lower Bollinger band: SMA – 2*standard deviation of stock prices

$$\text{Bollinger Band Percentage} = \frac{\text{Pricet} - \text{Lower Bollinger band}}{\text{upper Bollinger band} - \text{lower Bollinger band}}$$

Manual strategy (manual rule-based trader):

In this section, I introduce a manual strategy that utilizes the technical indicators and implements a set of rules to enter or exit positions in the stock. The manual strategy relies on past prices to determine if the stock price is oversold or overbought. The technical indicators that will be used for developing manual strategy are Momentum, price/SMA ratio and Bollinger band percentage.

The technique to determine a long or short position is based on the combination of two out of three indicators. As a result, if any two of the following conditions are met, the sell or buy signal would be created:

Long opportunity or buy signal:

1. Momentum > 0
2. Price/SMA >1.05
3. Bollinger Band % > 1

Short opportunity or sell signal:

1. Momentum < 0
2. Price/SMA < 0.95
3. Bollinger Band % < 0

These are the thresholds that were introduced in course presentation. I tried different thresholds around these values and these values gave me very good performance compared to others.

After determining the short and long signals based on the rules explained above, the trading would be started with allowable positions of 1000 shares long or 1000 shares short or 0 shares.

Assumptions:

Symbol: 'JPM'

Starting value = $100000

Commission = $9.95

Impact = $0.0

In-sample period = January 1, 2008, to December 31, 2009

Out-sample period = January 1, 2010, to December 31, 2011

Allowable positions: 1000 shares long, 1000 shares short, 0 shares
Lookback period for calculating indicators: 14 days
Benchmark: Buy 1000 shares on the first trading day and sell 1000 shares on the last day
The table below shows the comparison between manual strategy and benchmark for in-sample and out-sample periods.

	manual strategy		Benchmark	
	in-sample	out-sample	in-sample	out-sample
cumulative return	0.353	0.0229	0.0123	-0.0836
standard deviation	0.0125	0.0078	0.017	0.0085
mean	0.0007	0.0001	0.0002	-0.0001

The comparison between benchmark and manual strategy shows that for both in-sample and out-sample periods manual strategy is working better since it has higher cumulative return and average of daily returns. However, the comparison between in-sample and out-sample periods for manual strategy shows that in-sample period works extremely better than out-sample period.

I also plotted the normalized prices for Manual strategy and benchmark during in-sample period.

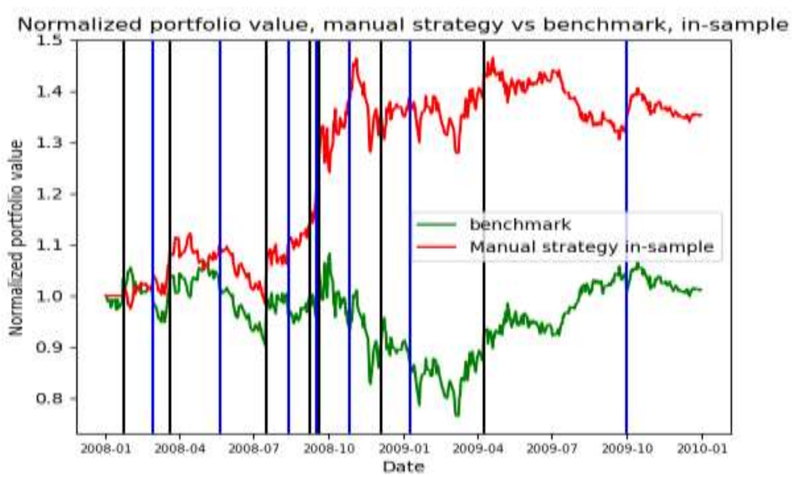


Figure 1

To compare the performance of manual strategy against benchmark, the normalized prices are plotted during out-sample period.

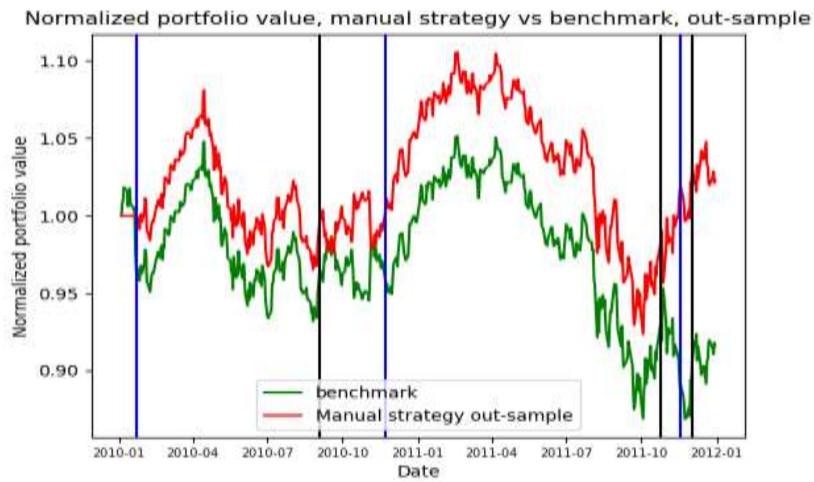
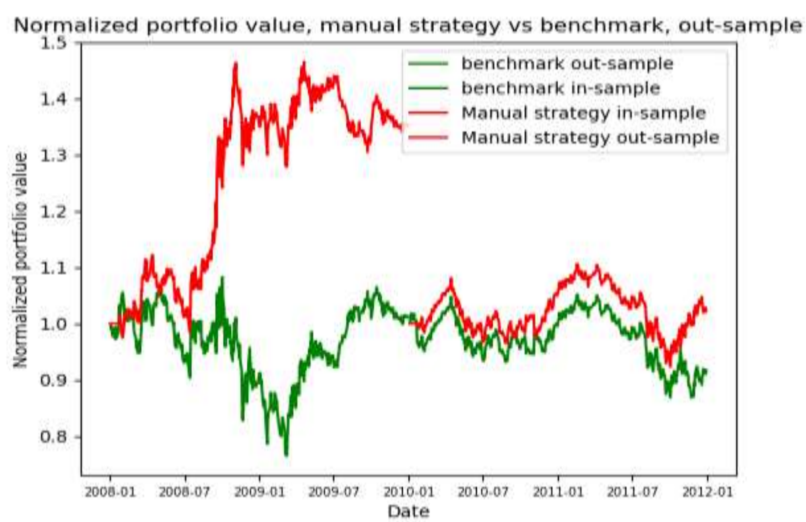


Figure 2

In both figure 1 and figure 2 during in-sample and out-sample periods, Manual strategy performs better than benchmark. However, in out-sample period manual strategy is closer to benchmark.

The following visualization shows the comparison between in-sample and out-sample periods for Manual strategy and Benchmark.



As we can see from this chart, the performance of manual strategy drops dramatically when moving from in-sample period to out-sample period. The reason is that the indicators' thresholds were optimized in a way that gives manual strategy the best performance during in-sample period, but this is not necessarily true for out-sample period.

Strategy learner:

In this part, I introduce a classification-based learner that can learn a trading policy using the bag learner (ensemble tree-based model) that utilizes RT learner (random tree learner) as the weak learner to create an ensemble algorithm. The features used to learn the policy are three indicators that were used to develop the manual learner: Momentum, Bollinger Band percentage and price/SMA ratio.

The actions for trading (buy, sell, do nothing) are converted to classes in the learning problem (+1, -1, 0). As a result, when the learner creates the class labels, it indicates an action in trading problem.

Setting up the learner:

The Strategy learner creates an instance of bag leaner with leaf size of 5 and bag size of 14, and the bag learner calls RT learner (Random tree) to create ensemble of trees.

Training step:

In the Add_evidence() method in Strategy learner, in-sample or training data is used to calculate features and labels. Three Technical indicators (Momentum, Bollinger Band % and price/SMA) are calculated and used as features to train the machine learning model.

To calculate the labels for each data point, the N-day(in this project 2-day) return is compared to the YSELL and YBUY to generate the train data with +1, -1, and 0 labels.

YBUY and YSELL are tuned to optimize the performance of the learner.

Testing step:

In the testPolicy() method, historical price data is used to calculate 3 Technical indicators(features). The query() method of the learner uses these features to generate Y test for the data.

After tweaking all the parameters of the strategy learner including (leaf size, bag size, YSELL, YBUY) the best possible Strategy learner for 'JPM' stock during in sample period (between 1st January 2008 and 31 December 2009) is generated with the following parameters:

Assumptions:

Symbol: 'JPM'
Starting value = $100000
Commission = $9.95
Impact = $0.0
In-sample period = January 1, 2008, to December 31, 2009
Out-sample period = January 1, 2010, to December 31, 2011
Allowable positions: 1000 shares long, 1000 shares short, 0 shares
Lookback period for calculating indicators: 14 days
Benchmark: Buy 1000 shares on the first trading day and sell 1000 shares on the last day
YSELL: 0.016
YBUY : -0.016
Leaf size= 5
Bag size =14
N (number of days for price comparison) = 2

Experiment 1:

In This experiment, I compared the performance of strategy learner (classification-based strategy) against manual strategy, and a benchmark. They all have been introduced in the previous sections.

Experiment steps:

1.Strategy learner:

After reading in the data, the add_evidence() method in strategy learner is used to train the leaner. Then the testPolicy() method uses query() method from bag learner to retrieve Y labels. These labels are used to trade stocks which results in the trade table. The portfolio value is calculated based on the trade table and after normalizing portfolio values they are displayed in comparison charts as green lines.

2. Manual strategy:

After reading in the data, testPolicy() method of manual strategy is called to generate trades table based on the rules explained in the previous sections, then the portfolio values and normalized portfolio values are calculated which are used to display the results of Manual Strategy in black color in comparison charts.

3.Benchmark:

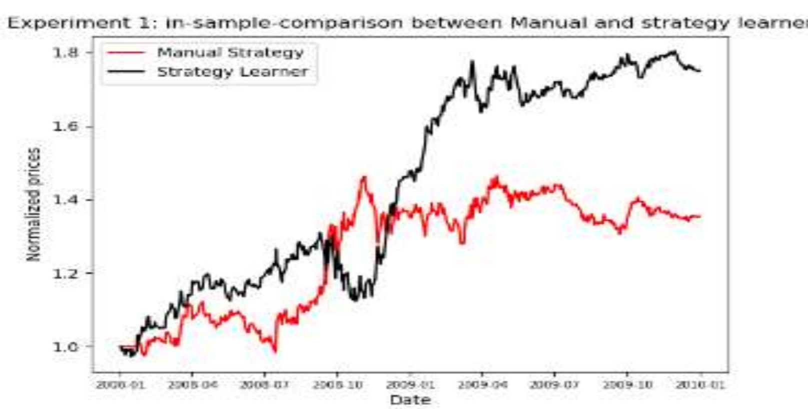
The benchmark is created based on some assumptions that we will explain in the assumption section, then the portfolio values and normalized portfolio values are calculated which are used to display the results in black color in comparison charts.

Assumptions:

Symbol: 'JPM'
Starting value = $100000
Commission = $9.95
Impact = $0.0
In-sample period = January 1, 2008, to December 31, 2009
Out-sample period = January 1, 2010, to December 31, 2011
Allowable positions: 1000 shares long, 1000 shares short, 0 shares
Lookback period for calculating indicators: 14 days
Benchmark: Buy 1000 shares on the first trading day and sell 1000 shares on the last day
Leaf size= 5
Bag size =14

Results:

The following visualization shows the results of experiment which is the comparison between the performance of strategy learner, manual strategy, and benchmark during in-sample period. During in-sample period both manual strategy and strategy learner outperform benchmark, and strategy learner outperforms manual strategy. In strategy learner we use a classification-based learner, and we tune the parameters in a way that performs the best during in-sample period. Also, previously we explained that the threshold for buy and sell in manual strategy was optimized to outperform benchmark during in-sample period, and this can be clearly seen in these visualizations.



After comparing the performance during in-sample periods, different strategies were compared against each other during out-sample period. The following visualization shows that strategy learner performs very poorly compared to manual strategy and benchmark during out-sample period. The reason is that I optimized the parameters of the bag learner in a way that produces the best performance during in-sample period, but this does not necessarily result in a good performance for out-sample.

statistics for strategy learner, manual strategy, and benchmark during in-sample and out-sample periods

	manual strategy		strategy learner		Benchmark	
	in-sample	out-sample	in-sample	out-sample	in-sample	out-sample
cumulative return	0.353	0.0229	0.7498	-0.9502	0.0123	-0.0836
standard deviation	0.0125	0.0078	0.0121	0.44	0.017	0.0085
mean	0.0007	0.0001	0.0012	0.0304	0.0002	-0.0001

Would you expect this relative result every time with in-sample date? Explain why or why not?

In most cases I expect strategy learner to outperform the benchmark and the manual strategy (even though in some small periods in the chart manual strategy outperforms strategy learner but for the most part it will beat manual strategy). Since we are using bag learner, which is an ensemble model, we would have some randomness in the learning process that would create the possibility of underperformance of strategy learner.

Experiment 2:

In this experiment, the strategy learner is used with different impact values to understand the effects of different impact values during in-sample period.

Hypothesis:

Impact is the cost that the buyer or seller of stocks occur when doing a transaction. If the impact is low, it would work to the favor of the trader since the trader doesn't have to pay extra money. The hypothesis here is that lower impact would cause the stocks perform better.

Hypothesis:

Parameters:

Symbol: 'JPM'
Starting value = $100000
Commission = $9.95
Impact = 0.05, 0.005, 0.0005

In-sample period = January 1, 2008, to December 31, 2009

Allowable positions: 1000 shares long, 1000 shares short, 0 shares
Lookback period for calculating indicators: 14 days
YSELL: - 0.016
YBUY : 0.016
Leaf size= 5
Bag size =14
The normalized portfolio values have been plotted for different impact values of 0.05, 0.005, and 0. 0005.As we can see in this plot the lower the impact value, the better the performance of the portfolio value would be.