Project 8: Strategy Evaluation

Evaluation of trading strategy using bagging learners with indicators

Ted Pham

Tpham328@gatech.edu

***Abstract*—**EMA9 crossover EMA21, MACD, and Bollinger bands are 3 indicators that are used to build a manual trading strategy to compared with a machine learning strategy that also use the same 3 indicators. Machine learning strategy proven to be very effective in stock trading but certain outside impact and more test needed to be considered.

INTRODUCTION

In our previous project, we have been exploring different indicators that can be useful to use in stock trading. The current paper discussed the test results of applying these indicators to back-test trading strategy. With the help of machine learning technique like bagging learners of classification trees, we compare the result using machine learning in trading using those indicators as significant features that signal buy, sell or hold cash with a manual trading strategy that commonly used among traders using those indicators.

The manual strategy is built based on how these indicators have been using in trading and combining these indicators to make trading decision. The three indicators using in these experiments are EMA9 cross EMA21, MACD, and Bollinger Band Percentage.

For simplicity, the in-sample data we are using to train the model is JPM stock prices from January 1, 2008 to December 31, 2009. Out of sample data is the same stock data from January 1, 2010 to December 31, 20211. The in-sample data is used to train the machine learning model, after that the model is not updated or learned with new data, only testing is done on the out-sample data. Starting cash is $100,000. And we can only trade JPM, our maximum holding at any time is 1,000 shares either in long or short position.

The machine learning we using here we called Strategy Learner is a bagging learns with bag size of 10 (iterate running 10 times) of random tree learners, a classification tree learners we have built from project 3. The random tree learns leaf size is 5.

It is popular belief that a combination of indicators should be used in trading decision and skeptical that machine learning can help in trading. The report indicates that using machine technique did improve trading performance using those indicators. The first experiment compared a manual strategy with machine learning strategy in term of performance of portfolio

Second experiment introduced certain load to the cost of buying or selling stock since no entry is perfect and some uncertainty increase your cost. This extra cost is considered as impact to the cost of the trade in addition to the commission fees.

1. INDICATOR OVERVIEW: (MANUAL STRATEGY)

a. INDICATOR 1: CROSSING OF EMA9 AND EMA21

The EMA9 cross EMA21 is a popular technical analysis tool used in stock trading that I personally have been using in the past. So the choosing of lookback window of 9 and 21 periods are personal experience.

To put EMA9 cross EMA 21 into a signal, I decided to use the difference between EMA9 and EMA21, a positive value indicates EMA9 is higher than EMA21, and indicating uptrend. A negative value of this difference indicates EMA9 is smaller than EMA21 and downtrend.

EMA9-EMA21 or EMA9x21 is a continuous variable that is used to train our Strategy Learner. The feature is used in random tree, bagging learners.

b. INDICATOR 2: MACD

MACD is used to identify trend changes, momentum and potential buy or sell signal when 2 lines cut. MACD is consists of 2 lines: The MACD line and the signal line. MACD line is the difference of the EMA 12 and EMA 26. The signal line is then calculated from the MACD line, it is the EMA9 of the MACD line .There are 2 lines in MACD so I used MACD histogram which is the difference between MACD line and signal line. Positive MACD histogram indicates bullish trend and negative MACD histogram indicates bearish trend.

We will use MACD histogram and EMA crossing as our main indicators to decide if the market is in up or down trend. And we trade follows the trend.

For Strategy learner, since MACD histogram is a continous variable, it is used to train our bagging learners in similar way as EMA crossing.

c. INDICATOR 3: Bollinger Band Percentage

Bollinger bands is used to measure the volatility of a stock price within a period. They consisted of 3 lines: The first line is a simple moving average (SMA), the top and bottom band is 2 standard deviation lines plotted from the middle line. To convert the Bollinger band into 1 simple Bollinger band percentage (BBP), when low BBP means stock price is closer to the bottom band and high BBP means stock price is closer to the top band.

Since there are 3 lines in Bollinger band, we opted to use BBP capture the position of stock price compared to the bands. A BBP of 50 or 50% indicates neutral, while the closer BBP to 100 or over 100, stock is considered overbought and might retrace back to 50. A BBP near 0 or below 0, indicate stock is oversold and BBP might retrace back to 50. However it is rare when this cross happens. After some test, I decided to use threshold of 70 and 30 to decide. When BBP over 70, stock price is considered overbought and high, it is unwise to entry long even though the trend might be still up. Vice versa, when BBP is below 30, it is considered oversold, it is unwise to short.

BBP is a continuous variable, and therefore can be used to train our Strategy learner. It is used as a feature in the bagging learner.

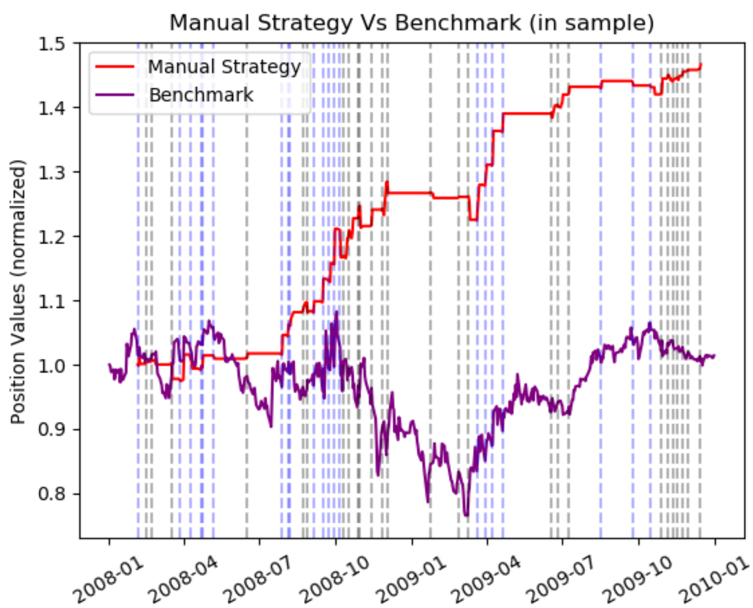
d. Manual Strategy.

The manual strategy we are using is “Follow the trend”, we will used EMA crossover and MACD histogram as our indicators to decide the trend, and trade with the trend. When EMA9x21 is positive and MACD histogram is positive, we only go long and look for entry to go long. In addition, we don’t want to entry when stock is overbought or high. So our entry for long is when all these conditions are met. EMA9x21 is positive, MACD histogram is positive, and BBP is not over 70. In our program we considered this as vote = 1 or long signal.

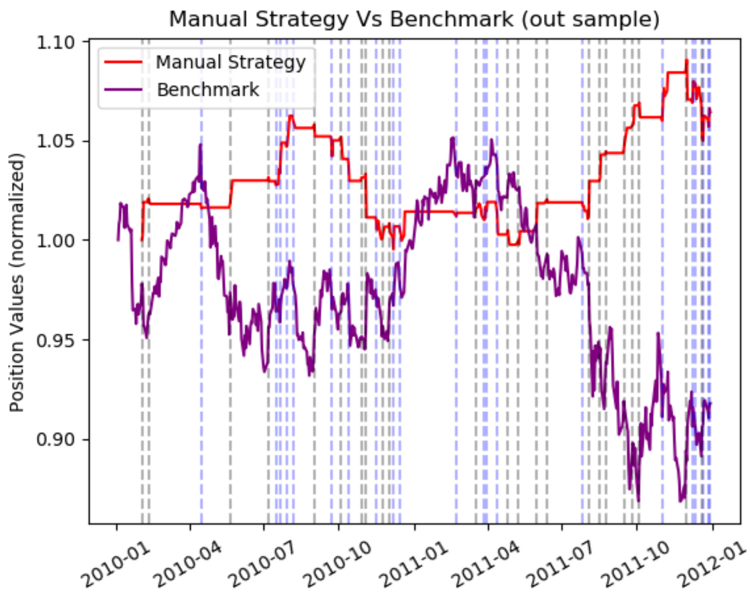
Vice versa, we want to enter short when trend is bearish but only enter when stock is not oversold. When EMA9x21 is negative, MACD histogram is negative, and BBP is not under 30 (oversold) we will enter short position. In our program we considered this as vote = -1 or short signal.

When any of this condition is not met, we considered this as neutral position and will get out to hold cash. For example, when we are currently in long position and BBP is over 70 or overbought, we will sell and exit. If BBP is not over 70 yet, but either MACD or EMA9x21 indicate trend change (negative value) we also exit our long position.

We did cheated a little bit here because we have looked at the in-sample data and confirm that our indicators work well. Here is our result if we trade the stocks during in sample period (2008-2009):



.Here is the result when we use the same manual trading strategy during out-sample period (2010-2011). Of course, we did not “peak”:



As you can see, our manual strategy beats the benchmark convincingly. The chart supported that our manual strategy results in higher portfolio value in both in-sample and out-sample test case.

The benchmark is buy and hold 1,000 shares of JPM until end of period. In out-of-sample test, we can see that our performance drop significantly from 46% gain to only 7% gain. This is understandable because our biased looking into the future during in-sample period test. The table below summary our Manual Strategy using in-sample and out-sample data.

		Manual Strategy	**Benchmark**
In Sample	Cumulative Return	0.4666814	0.01421
	Average daily return	0.00083646	0.000172
	Standard Deviation daily return	0.0065675	0.0170416
	Sharpe Ratio	2.02183	0.1606
	Cumulative Return	0.06424	-0.08204

		Manual Strategy	Benchmark
Out Sample	Average daily return	0.0001334	-0.00013
	Standard Deviation daily return	0.00302	0.00849
	Sharpe Ratio	0.69972	-0.25047

2. STRATEGY LEARNER

We have 3 indicators that are both continuous numbers, however, our target variable is signal when we should buy or sell the stock. This indicate we should use a classification machine learning approach. Two of them come to mind: linear classification or decision tree, and decision tree makes more sense here, because it can take in regression and continuous features as independent variables while predict classification target variable (buy, sell and hold).

We have the indicators values calculated from the data. So we only need to calculate the target variable. To convert the price of stock into decision buy or sell, we decide to roll-forward the data. Using a future period of 10 days, if stock price in 10 days is 2% higher than the current price, then we considered this a buy target or our target variable is 'BUY'. Vice versa, if it is 2% lower, target variable y is 'SELL'. If it is between -2% and 2% then it is not worth the risk so y is neutral or 'HOLD'.

We have run some sensitivity tests to decide the 10 days period and 2% threshold here. Given our data span is only 2 years and impact is 0.5%, these parameter is appropriate.

Using bagging learners with random trees for our strategy learner here is also appropriate because:

- Bagging improve accuracy with each bag is run on a different tree learner.
- Bagging reduce overfitting
- Bagging with random tree learner can handle non-linear relationship (buy, sell, hold signals)

- Bagging allows feature and provide insights to features importance, features are indicators here.

I did not have to discretized the data because I didn't use Qlearner, our indicators are calculated from stock price and are continuous so there is no need to standardized when using to train our Strategy Learners. We did have to remove the first 20 days in the data because BBP is not available until the 20^{th} day, as well as remove the last 10 days or "peak into future" period to calculate the target variable y.

We choose our bag size of 10 so it is not taking too long to run, and our leaves size of 5 seem to be good choice, lower than that might cause overfitting.

For this part of the project, we use the JPM data from Jan 1, 2008 to Dec 31, 2009 to train the learners and test. We also considered the impact of 0.5% and commission of $9.95. Chart below show showed the result using in-sample data to test

Although not being asked by the assignment, we did test the learner using out-sample data from Jan 1, 2010 to Dec 31, 2011 and have a good result as well.

3. EXPERIMENT1: COMPARE MANUAL VS STRATEGY

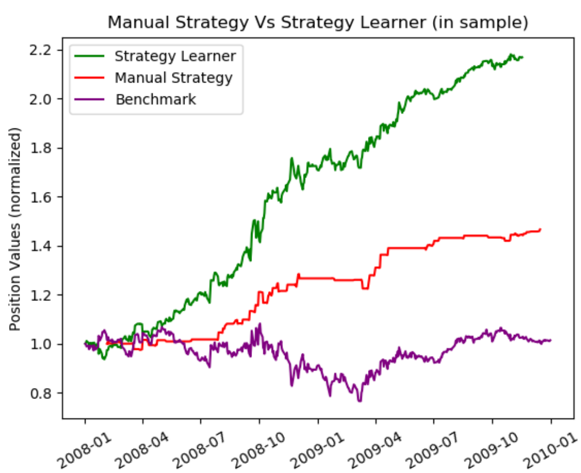
We use the same in-sample and out-sample data of JPM stock to conduct our manual strategy with learners strategy and compare the results.

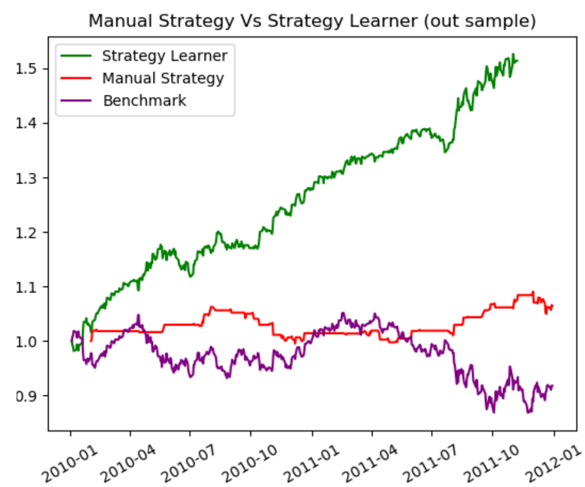
Our starting cash is $100,000. Our limit is 1,000 shares at any time, meaning we can only hold 1,000 shares in either long or short position at any time. Comission is assumed to be $9.95 and impact is 0.5%.

For this experiment we used the same Manual Trading Strategy as described in section 1.D.

We used the same Strategy Learners described in section 2 of the report. Here we trained the data using in-sample data, and test the data using out-sample data. We will not let the learners learn any new data from out-sample data. Strategy Learners is bagging learner with bag size of 10 and leaf size of 5.

The benchmark is buy 1,000 shares of JPM from the beginning and hold until the end of the period. Here is the result of data:





In both cases, our Strategy Learner outperformed both the Manual Strategy and the benchmark significantly. This indicates that using machine learning did help with trading. Our Manual Strategy also outperformed the benchmark so this is a good strategy as well but no where near our Strategy Learner.

In out-of-sample test, our Strategy Learner also outperformed Manual Strategy significantly. Also, the percentage gain has been dropped (from over 116% gain to 51%). We expected this outcome since future is unpredictable.

In in-sample test, Strategy Learner beats our Manual Strategy. We expected this outcome as well since Manual Strategy using indicator is not fine tune like machine learning, and indicators calculated from historical data is considered a lagging compared to price so it cannot be performed as good as Strategy Learner, but it is still a good strategy to use. You might see more things from the indicators rather than the Strategy Learners, which only use vector indicators.

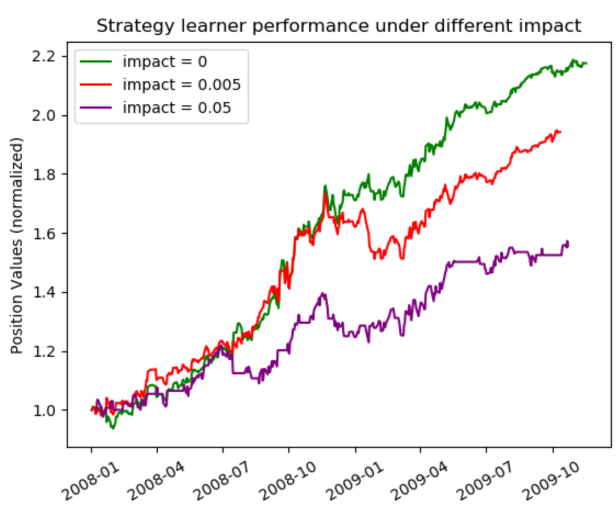
4. EXPERIMENT2: CONSIDER THE IMPACT

Changing the impact reduce the performance of our Strategy Learner significantly. Impact is the extra cost that we have when we enter a trade. By default impact value is only 0.5% of the price. If the impact is high, not many trade is profitable as we already in lost position when we entered trade, and only gain if the stock outperform the impact value.

When we built the Strategy Learner, we assume that we know the future and train the data using target variable y that we calculate by look into the future n = 10 days. If the price in 10 days is 2% higher than today, we enter trade and vice versa. With impact of 0.5%, even if we entered trade at 0.5% higher cost, in 10 days, if stock price is 2% or higher than current price, we still net a gain. Therefore y = 'BUY' is still correct decision, but what if impact = 2.5%. In 10 days, if stock price only gain 2% while your cost is 2.5% higher than current price, we net a loss in 10days, therefore, y = "BUY" is incorrect decision.

Impact directly affect the target variable y and therefore, it have significant impact with Strategy Learner.

To illustrate our hypothesis, we run a sensitivity test using impact = 0, 0.5% and 5%. The result of our portfolio using Strategy Learners under different impact values follows:



The higher the impact, the less profitable or worse performance our Strategy Learner performed. We also calculate the following metrics to compare:

	Cumulative Return	**Average daily return**	**Standard deviation daily return**	**Sharpe Ratio**
Impact = 0	1.1751	0.00169	0.01090	2.4745
Impact = 0.5%	0.9423	0.00154	0.01093	2.243
Impact = 5%	0.55525	0.001048	0.01214	1.370
Result of increase impact to the metrics	Decrease, decease in profit	Decrease	No significant change	Decrease

CONCLUSION

EMA 9 cross over EMA21, MACD, and Bollinger Bands are all very good indicators use in trading. The manual strategy built based on these 3 indicators produced a great result. Using these 3 indicators as features in a machine learning model like what we built in Strategy Learners produced outstanding result, although more test needed to be considered. Our test on JPM stock is also suffer from survivor bias as we know the stocks survive in SP500. Regardless, machine learning are proving to be very effective in stock trading although more things need to be considered rather than just buy and sell signal like the actual entry price or commission fees.

REFERENCES:

https://school.stockcharts.com/doku.php?id=trading_strategies

https://school.stockcharts.com/doku.php?id=technical_indicators:stochastic_oscillator_fast_slow_and_full

My own Project 6: Indicator Evaluation report