Project 8:
Strategy Evaluation

Liam O'Donnell
lodonnell32@gatech.edu

Abstract—CS 7646: Machine Learning for Trading was composed of 7 projects that laid the groundwork for Project 8: *Strategy Evaluation*, the course's capstone project. The project used the student's previous implementation of Random Forests (built in Project 4), along with select technical indicators (Project 6) to build a manual strategy and a strategy learner.

1 Indicator overview

The student researched five (5) indicators in Project 6. For project 8, both the manual strategy and strategy learner utilized three (3) of the indicators: Bollinger Bands, RSI, and EMA.

1.1 Bollinger Bands

The implementation of Bollinger Bands is as follows: for a given stock and day, calculate the 14 day moving average and rolling standard deviation. The upper Bollinger Band is defined as the moving average for a given day + the rolling standard deviation. The lower Bollinger Band is defined as the moving average - the rolling standard deviation. A "bullish" outlook is when the price of a stock passes the lower Bollinger Band. Alternatively, a "bearish" outlook is when the price of a stock passes the upper Bollinger Band. Both the strategy learner and the manual strategy use a 14 day moving average and standard deviation as parameters. The manual strategy uses the upper and lower band. The strategy learner uses both the upper band and lower band as well as the distance of the stock's price to the upper and lower bands.

1.2 Relative Strength Index (RSI)

The relative strength index (RSI) is a momentum-based indicator. It is a number between 0-100 indicating if a stock is overbought or oversold. If the RSI is > 70, the stock is considered overbought (bearish). Alternatively, if the RSI is < 30, the stock is considered oversold (bullish). The RSI is calculated by first subtracting the stock's adjusted close on each trading day from the previous day. If that value is > 0, it is a "gain", otherwise, if the value is < 0, it is a loss. The relative strength is calculated by dividing the rolling average gain (the student used a lookback period of 14) by the rolling average loss.. The RSI is then calculated by the formula below.

$$RSI = 100 - \frac{100}{(1 + relative\ strength)}$$

For the manual strategy and strategy learner, a lookback period of 14 days was applied. The manual strategy used the RSI at face value for each day in the trading period. The strategy learner used both the RSI and a separate value "new RSI" which subtracted 50 from the RSI value.

1.3 Exponential Moving Average (EMA)

The Exponential Moving Average, while directly related with a simple moving average, places a larger weight on more recent data points (relative to the day being observed). The EMA uses a smoothing parameter (α) to determine the weight to put on more recent days. The student generated his own, unique approach to implementing the EMA. When the EMA crossed above the SMA, a bullish signal was generated, alternatively, if the EMA crossed below the SMA, a bearish signal was generated. The manual strategy uses the EMA and the previous day's SMA to generate a signal, which the strategy learner uses the difference between the EMA and SMA, the delta between the price and SMA and the difference between the price and SMA on the previous day.

2 Manual Strategy

The first stage of creating a strategy was creating a manual strategy. The student defined the actions to take based on the technical indicators discussed above. The buy/sell actions, along with results of the manual strategy are discussed below.

2.1 Implementation

The manual strategy combines the data used in all three indicators into one pandas dataframe, indexed by date. A corresponding action column is created for each indicator (Bollinger action, RSI action, EMA action). If a bullish or bearish signal is generated from a given signal, the corresponding action column is populated with a 1 (buy signal) or a -1 (sell signal). If there is no bull/bear signal given, the respective action column is null.

Once the action columns are populated, the student assigned corresponding weights to each action. A higher weight is associated with a higher "importance" given to that indicator. Additionally a threshold was established. If the sum of the weights on a given day was greater than or equal to the threshold, an action was taken. The optimal weights matrix and threshold were determined through an extensive cross-validation process, where the weights for each of the five indicators from Project 6, along with the overall threshold, were tested to maximize returns across a diverse set of stocks. This process led to the student's selection of three specific indicators, as discussed in the 'Indicator Overview' section. The final weights following cross-validation were as follows: (Bollinger action: 0.5), (RSI action: 0.5), (EMA action: 1.5). The final threshold following cross-validation was 2.

After assigning weights and setting the threshold, each day was classified with a bullish, bearish, or neutral outlook. For example, if the Bollinger action was -1 (indicating a sell signal), the RSI action was 0 (hold signal), and the EMA action was -1 (sell signal), the weighted sum would exceed the threshold of -2 by applying each action's respective weight: (-1 * 0.5 + 0 * 0.5 + -1 * 1.5) = -2. This logic was applied daily, generating a complete outlook column indicating whether the manual strategy had a bullish, bearish, or neutral outlook.

The final step in the process was generating trades. An order was generated if a given day had a bullish/bearish outlook while respecting the project's constraints on shares owned (+1000 or -1000 shares).

2.2 In-Sample and Out-of-Sample Results

The manual strategy was refined using in-sample data and tested using out of sample data. The in sample data contained stock price and indicator information for the stock "JPM" from January 1st, 2008 to December 31st, 2009. The out-of-sample data contained stock price and indicator information for the same stock from January 1st, 2010 to December 31st, 2011. The in-sample cumulative return results are shown in **Figure 1** and the out-of-sample cumulative return results are shown in **Figure 2**. The benchmark shown is starting with the same amount of capital ($100,000), buying 1000 shares of JPM on day 0 and holding. One additional important note was that if a buy signal was given on day x, the trade was executed on day x+1. This was put in place to simulate real market conditions.

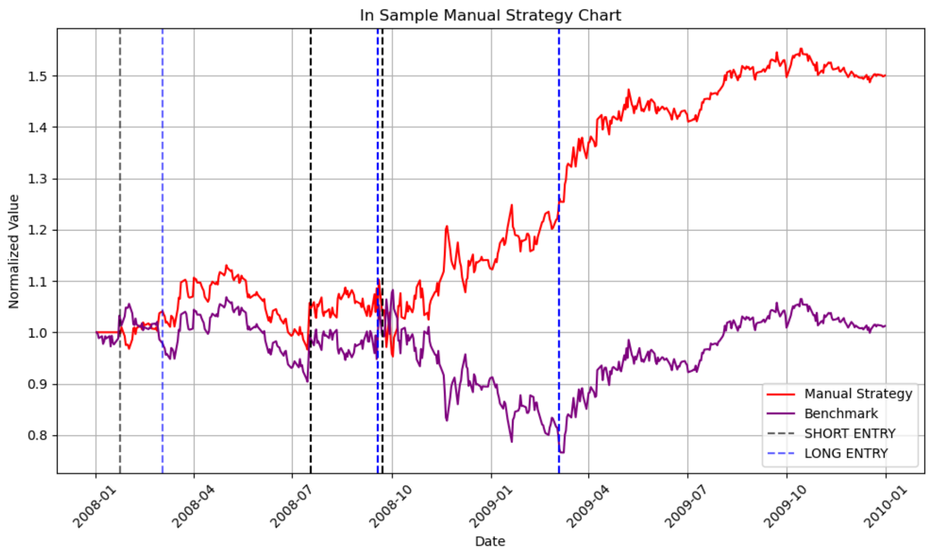

Figure 1—In Sample Manual Strategy

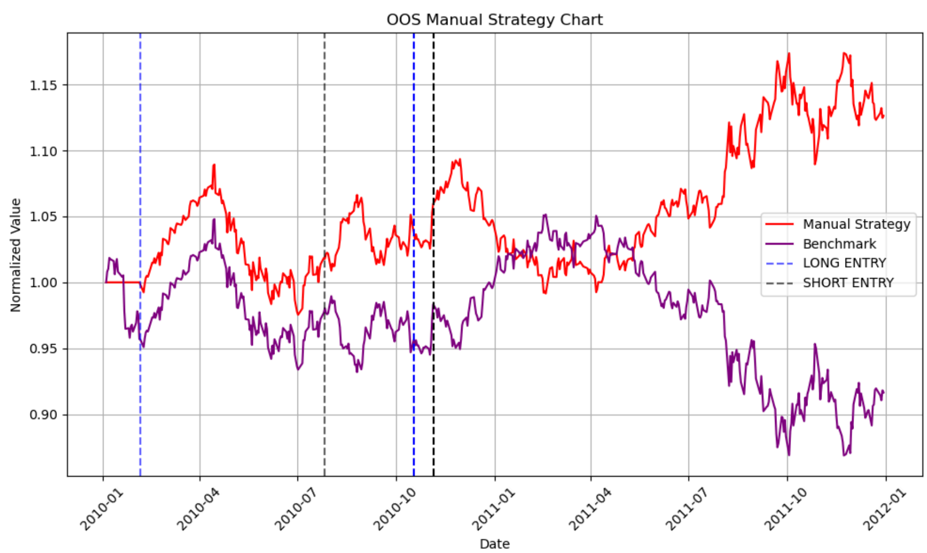

Figure 2—Out of Sample Manual Strategy

The student was optimistic about both the in-sample and out-of-sample results. The in-sample results beat the given benchmark by approximately 50%. The out-of-sample results beat the given benchmark by approximately 20%. The student expected the in-sample results to be "more optimal" when compared to the out of sample results due to the "human-fitting" of the model. The student determined the best weights and threshold to use that maximized return over the in-sample period on a specific stock. While the results were favorable in this use-case, the weights for another out-of-sample set of stock price data with an alternative symbol could exhibit a potentially over-fit manual strategy to the in-sample period and stock. For future exploration, the student may explore the manual strategy's effectiveness on stock symbols outside of "JPM".

Additional metrics for the manual strategy's in-sample and out-of-sample's testing are shown in **Table 1** and **Table 2**. As previously discussed, the in-sample and out of sample results were more favorable with regards to Sharpe ratio, standard deviation, cumulative return and mean of daily returns. The high Sharpe ratio in both the manual strategy and strategy learner when compared to their respective benchmark indicates a well performing, less volatile strategy.

Metric	Benchmark	Manual Strategy
Cumulative Return	1.23%	50.02%
STDEV of Daily Returns	1.7%	1.4%
Mean of Daily Returns	0.02%	0.09%
Sharpe Ratio	0.16	1.02

Table 1—In Sample Manual Strategy Performance Metrics

Metric	Benchmark	Manual Strategy
Cumulative Return	-8.36%	12.63%
STDEV of Daily Returns	0.85%	0.73%
Mean of Daily Returns	0.01%	0.03%
Sharpe Ratio	-.26	0.57

Table 2—Out of Sample Manual Strategy Performance Metrics

2 Strategy learner

Following the implementation of the manual strategy, the student built a machine learning strategy learning using bagged random trees. The implementation and results are discussed below.

2.1 Implementation

The student used the same indicators discussed above. One of the key challenges of the strategy learner was supplying the feature data in such a matter that would ultimately lead the trading bot to make actionable buy/sell decisions. As discussed in the "Indicators" section of the report, the following fields were derived from the Bollinger Band, RSI and EMA fields: Upper band, lower band, SMA, upper_distance, lower_distance, SMA_EMA_diff, price_to_sma, sma_ema_lag, and new_rsi. Additional parameters the student needed to consider when forming the training process were as follows: Limit (the % change significant enough to classify that day as a buy, sell, or hold), look_ahead (the number of days to "look ahead" to calculate the percent difference between the current share price and the "look ahead" day), bags (the number of random trees within the random forest), and leaves (the number of samples within a node to determine if the node was a leaf). To select the best parameters and features, the student undertook multiple rounds of grid searching. The first round was conducted out of add evidence and produced the most optimal feature set with regards to cumulative return to train the model. This step additionally produced a variety of additional parameters which helped the student narrow down which model parameters to use in the second step of grid search.

Once the student had selected his feature list, the proceeding grid search occurred within the training of the model itself. This was performed by supplying a list of limits, look_ahead values, bags, and leafs and having the model query itself to determine the optimal parameters in order to maximize cumulative return on the in-sample data. Rather than defining a set of parameters based on a signal, this approach sets parameters based on *supplied training data*. This step was approved via Ed Discussion. The student explored alternative, standardized metrics such as Bollinger Band Percent and a scaled difference of the EMA subtracted by the SMA of a previous day, as well as the RSI score but this returned returns over the in-sample data that were lower than the non-scaled column set. A final key component of the student's strategy was accounting for randomness. The student back tested the strategy learner against several random seeds and continued to get favorable in-sample results.

When the model was queried, it used the mode of the returned results to determine the long,short or hold outlook for the given day. Trades were executed in the same manner discussed in manual strategy. Because of the extensive hyperparameter testing conducted, the student was able to assemble both the training and testing steps to execute within the given time constraints.

2.2 In-Sample and Out-of-Sample Results

Grid searching in the training phase proved to be largely beneficial as it relates to in-sample data, but overfitting evidence arose when applied to out of sample data. This is due to the optimized parameters for the supplied training data. While a majority of supplied training sets may use the same look ahead value, the model created may be fit too specifically to those instances of training sets. To increase model performance as it relates to out of sample performance, it would be best to train the model over multiple sets of concatenated training data. That way the non-standardized features could potentially be better generalized. The overfitting on in-sample data is shown in the following section.

3 EXPERIMENT 1

Experiment 1 was designed to analyze both in-sample and out-of-sample performance for the strategy learner and manual learner. Each learner trained and evaluated against the same in-sample benchmark (JPM from January 1, 2008, to December 31, 2009, as discussed in Section 1) and tested on the same out-of-sample benchmark (JPM from January 1, 2010, to December 31, 2011). The student hypothesized that while the strategy learner would outperform both the benchmark and the manual strategy in-sample, it might suffer from overfitting, potentially leading to lower performance compared to the manual strategy in out-of-sample data. The experiment was conducted with a starting cash balance of $100,000, an impact cost of 0.005, and a commission of $9.95. The strategy learner was trained on in-sample data and then evaluated on both in-sample and out-of-sample datasets, while the manual strategy applied the aforementioned, predefined rules to both datasets. The results for experiment 1 are presented in **Figure 3** and **Figure 4**. As shown in the figures, the student's hypothesis was confirmed. The strategy learner performed optimally when tested over in-sample data. While the strategy learner still beat the benchmark in the out of sample case, its slope indicated a certain level of overfitting to the in sample data. The manual strategy, as discussed in its respective section, proved to perform well on both the in sample and out of sample results.

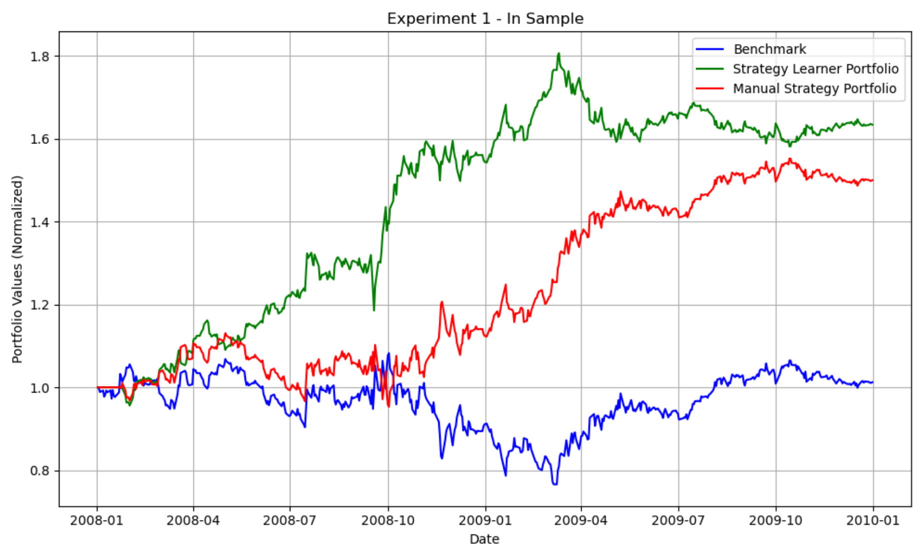

Figure 3—Experiment 1: In-Sample Learners vs. Benchmark

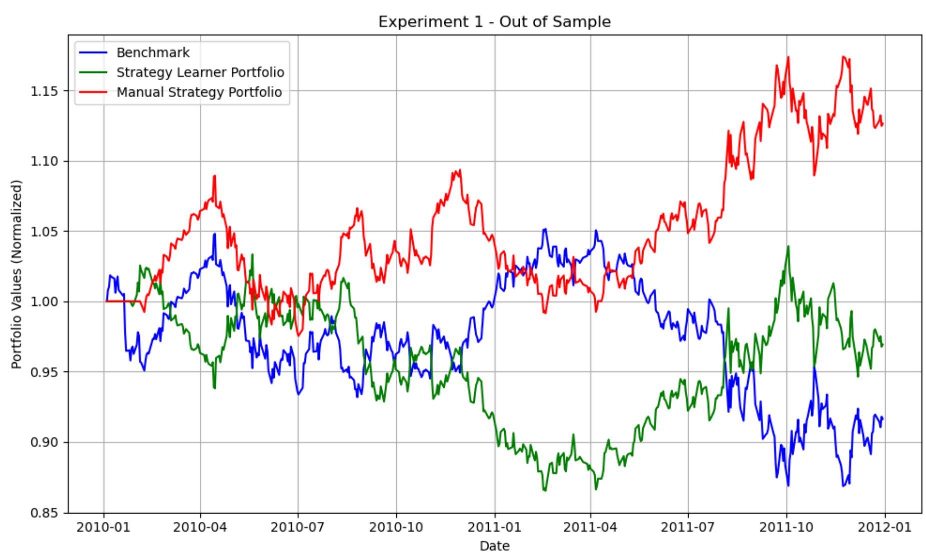

Figure 4—Experiment 1: Out-of-Sample Learners vs. Benchmark

3 EXPERIMENT 2

Impact is defined in Project 5: Marketsim as "the amount the price moves against the trader compared to the historical data at each transaction". Experiment 2 shows how the changing of the impact value affects sample trading behavior for the Strategy Learner. The experiment was created to test the student's strategy learner over several impact values, and show how it affected at least two metrics (chosen by the student) for the strategy learner. The following impact values were tested: 0, 0.1%, 0.5%, 1%, 4%. The student implemented his solution by looping through each impact value, training over the in-sample data, then querying the generated model with the same, in-sample data. Cumulative returns and cumulative trades were tracked during the loop and plotted for comparison. The results for experiment 2 are shown in **Figure 5** and **Figure 6**. The student hypothesized that decreasing the impact would lead to both higher returns and a higher value for cumulative trades. This student arrived at this hypothesis because a lower impact would provide a price closer to the "actual" price (the price used to determine whether the outlook was bearish or bullish, and thus, determine the action to take). While the student's hypothesis was confirmed for impacts 4% and 1%, the student's hypothesis was rejected for impacts 0.5%, 0.1% and 0%. An impact of 0% led to higher returns, but a lower cumulative impact when compared to an impact of 0.01%. The student believes that this could have been affected by one of two things, either the potential randomness as a result of the seed, or the 0.1% impact working in favor of the strategy learner. In future experiments, the student could test both measures against multiple signals and multiple seeds in order to analyze if this behavior persists.

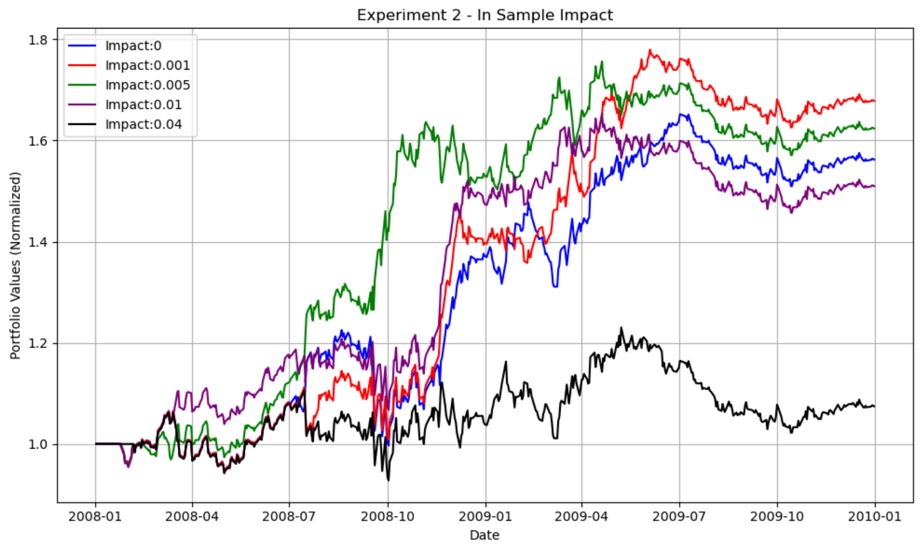

Figure 5 —Experiment 2: In-Sample Impact Comparison (Portfolio Value)

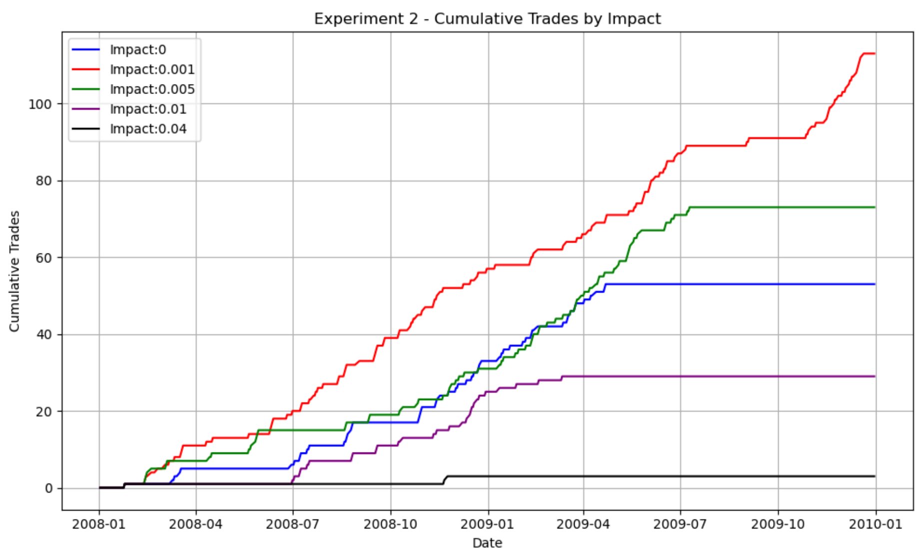

Figure 5 —Experiment 2: In-Sample Impact Comparison (Cumulative Trades)

4 REFERENCES

1. Bollinger, J. (2002). Bollinger on Bollinger bands. McGraw-Hill.
2. Chen, J. (2024, April 6). What is EMA? How to Use Exponential Moving Average With Formula. Investopedia. https://www.investopedia.com/terms/e/ema.asp#:~:text=An%20exponential%20moving%20average%20(EMA)%20is%20a%20type%20of%20moving,the%20exponentially%20weighted%20moving%20average.
3. Fernando, J. (2024, September 23). Relative Strength Index (RSI) indicator explained with formula. Investopedia. https://www.investopedia.com/terms/r/rsi.asp#:~:text=The%20relative%20strength%20index%20(RSI)%20is%20a%20momentum%20indicator%20used,the%20price%20of%20that%20security.
4. Hayes, A. (2024, September 6). Stochastic Oscillator: What it is, how it works, how to calculate. Investopedia. https://www.investopedia.com/terms/s/stochasticoscillator.asp
5. Mak, D. K. (2021). Exponential Moving Average. In Trading Tactics in the Financial Market. Springer International Publishing AG.
6. Ni, Y., Liao, Y.-C., & Huang, P. (2015). Momentum in the Chinese Stock Market: Evidence from Stochastic Oscillator Indicators. Emerging Markets Finance & Trade, 51(sup1), S99–S110. https://doi.org/10.1080/1540496X.2014.998916
7. O'Donnell, L. (2024). *Project 6: Indicator Evaluation* [Unpublished]. Georgia Technological Institute
8. Schlossberg, B. (2024, July 27). How to trade the MACD. Investopedia. https://www.investopedia.com/articles/forex/05/macddiverge.asp
9. Thompson, C. (2024, April 16). Bollinger Bands: What they are, and what they tell investors. Investopedia. https://www.investopedia.com/terms/b/bollingerbands.asp
10. Thomsett, M. C. (2019). Relative Strength Index. In Understanding Momentum in Investment Technical Analysis. Business Expert Press.
11. Wang, J., & Kim, J. (2018). Predicting Stock Price Trend Using MACD Optimized by Historical Volatility. Mathematical Problems in Engineering, 2018(2018), 1–12. https://doi.org/10.1155/2018/9280590