

Project 8: Strategy Evaluation

Richard Albright

903548616

ralbright7@gatech.edu

Abstract—The development of a Q-learning trading strategy in comparison to execution of both a rules based trading strategy and a buy and hold benchmark.

1 INTRODUCTION

The project develops a Q-learning trading strategy that is compared to both a manual rules based, and buy and hold trading strategies. Five indicators were used for both the manual rules based and Q-learning trading strategies. The indicators used for both the Q-learner and the rules based strategy are; simple moving average (SMA), exponential moving average (EMA), bollinger band percent (BB), moving average convergence divergence (MACD), and momentum (MOM). Both in and out of sample periods are analyzed for all 3 strategies.

2 INDICATOR OVERVIEW

The SMA used a lookback period of 50 days. The EMA used a lookback period of 10 days. Since the EMA is weighted more heavily to more recent price action than the SMA, the EMA oscillates around the SMA. The EMA and SMA values were then normalized. The normalized SMA was subtracted from the normalized EMA to create a crossing (CROSS) indicator to reduce the dimensionality of indicators from 5 to 4. BB used a lookback of 40 days, and smooths the data using a 3 day exponential moving average. The bollinger band percentages were set at levels above and below 1 standard deviation away from the mean. MACD used the industry default values of a 12 day lookback for the fast average, a 26 day lookback for the slow average, and a 9 day lookback for the actual MACD signal. MOM used a 3 day lookback. Both the Manual Strategy and the Q-learner used the same parameters for all indicators.

3 MANUAL STRATEGY

The manual rules based strategy was created around the idea of mean reversion of the indicators. The author used to implement a long/short equity pairs trading strategy for a proprietary trading firm, which focused primarily on re-

version to a mean ratio. Prior experience was used to attempt to capture mean reversion in a single equity using the indicators above. The tuning of lookback periods and smoothing variables were a result of trial and error while tuning on the in-sample period.

A long or short signal was then generated using a boolean value for every trading day to determine if the holdings should be long or short the stock. A long signal for CROSS and MACD consisted of the value crossing 0 on the upside and not being long the stock, while the short signal crossed 0 on the downside while not being short. The long signal for BB was crossing above 0.8 on the upside and not being long the stock, while the short signal was crossing below -0.8 on the downside while not being short the stock. MOM was used to prevent from exiting a prior signal too early as a long MOM signal was crossing 0 on the upside while being long the stock, and a short MOM signal was crossing 0 on the downside while being short the stock. All signals above either are mean reverting or oscillate around 0, providing multiple opportunities to enter and exit a long or short position. As long as the price of the underlying security is stable, and there is sufficient volatility, it should generate profitable trading opportunities using mean reversion. Therefore, the signals derived from the indicators above were combined using an or statement on all signals to create a single long (1) or short (-1) signal for every day during the period.

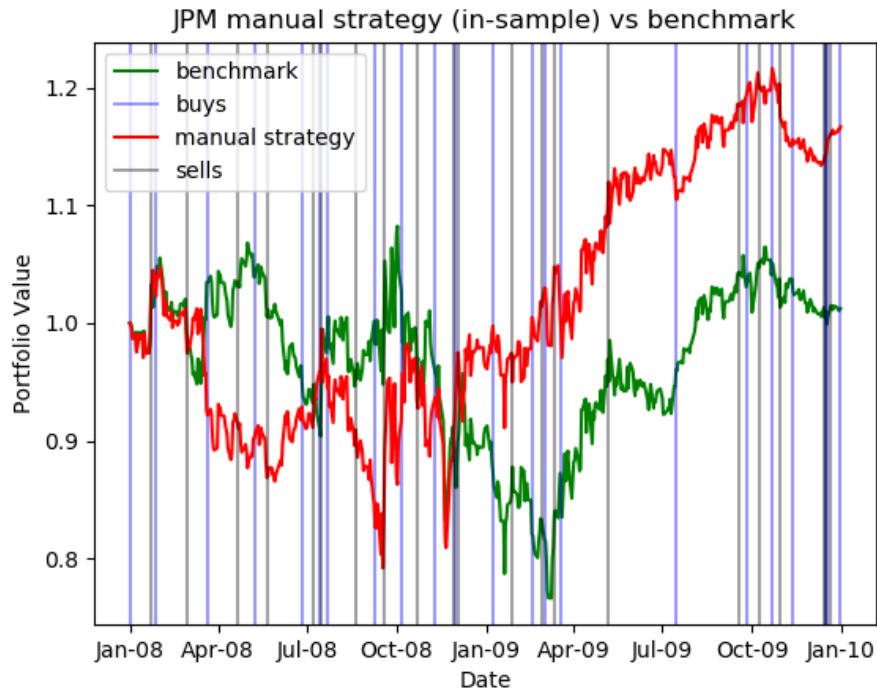


Figure 1

Figure 1 represents the results of the execution of the manual rules based trading strategy trading JPM during the in sample period between January 2008 through December 2009. Starting with an initial cash outlay of \$100,000, the manual strategy had an ending value of \$116,703, while the benchmark had an ending value of \$101,230.

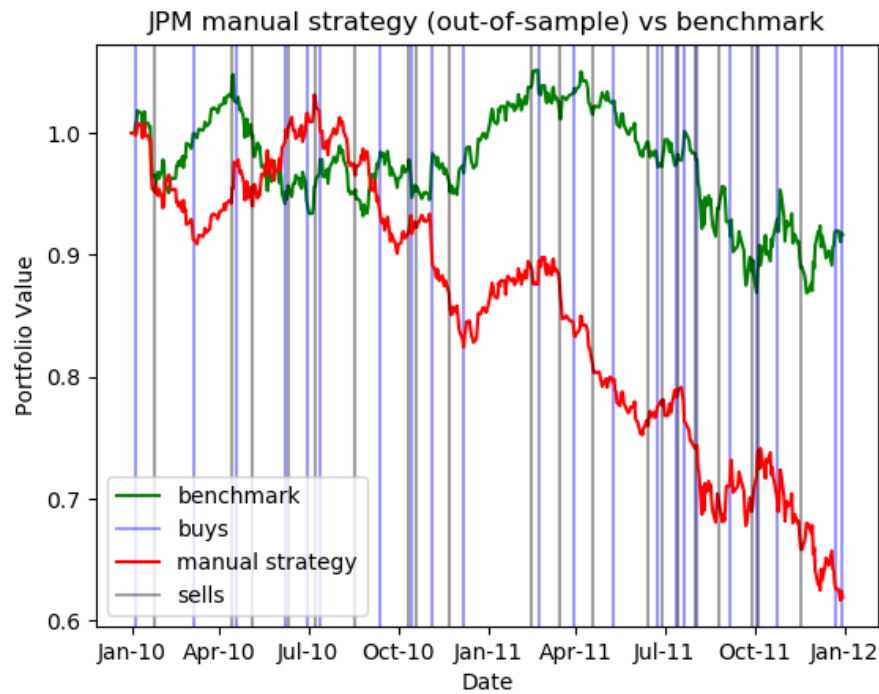


Figure 2

Figure 2 represents the results of the execution of the manual rules based trading strategy trading JPM during the out of sample period between January 2010 through December 2011. Starting with an initial cash outlay of \$100,000, the manual strategy had an ending value of \$61,936, while the benchmark had an ending value of \$91,660.

Below is a table of summary metrics of the Manual Strategy vs the benchmark for both the in-sample and out-of-sample periods, including commissions and slippage (impact).

| Metric | Manual (in) | Bench (in) | Manual (out) | Bench (out) |
|------------------|-------------|------------|--------------|-------------|
| Cum Return | 0.167 | 0.0123 | -0.3806 | -0.0834 |
| Mean Daily Ret | 0.0004 | 0.0002 | -0.0009 | -0.0001 |
| STD of Daily Ret | 0.0042 | 0.0141 | 0.01 | 0.0083 |
| Sharpe Ratio | 0.4148 | 0.1542 | -1.3816 | -0.2526 |
| Ending Value | \$116,703 | \$101,230 | \$61,936 | \$91,660 |

Table 1—Manual Strategy vs Benchmark Metrics

Clearly, the Manual Strategy for the in-sample period is overfit. The strategy performs worse than the benchmark in the out of sample period. The average standard deviation or daily returns is also lower at 0.0083, vs 0.0141 during the in sample period. The mean daily return is also lower at -0.001 for the out of sample period, while higher at 0.002 during the in sample period. Besides being overfit to the in sample period, the manual strategy likely requires more volatility to be present in order to be profitable.

4 Q-LEARNING STRATEGY

A Q-learner was implemented using the same indicator parameters as the manual strategy. Instead of combining the indicators using mean reversion signals, a Q-learner table was used to determine when to enter a long or short position based upon using the post 1 day return as a reward. The idea being to peek ahead one day into the future to determine which action to take. The allowed actions are Sell (0), Hold (1), and Buy(2). This is determined by looking at the current state and determining an action to take based on the reward. The state of the Q-learner was determined by splitting the CROSS, BB, MACD, and MOM signals in half along the median and discretizing by rank * 1000 for CROSS, 100 for BB, 10 for MACD, and 1 for MOM. The parameters were as follows for setting the state during the learning phase. The number of states was 1112. The number of actions is 3, alpha is 0.2, gamma is 0.9, random action rate is 0.3, random action decay rate is 0.9, and dyna was set to 0. Default values were used the parameters alpha through dyna, with the exception of the random action rate, which was lowered to 0.3 in order to speed up learning. The maximum number

of epochs to iterate over was set to 1000 with convergence being defined as > 3 cumulative returns being within 0.001 of each other in successive iterations. After learning over the in sample period, trades were then generated looking up the state based upon the discretized indicator values, for both in and out of sample periods. A random seed was set to the author's GTID.

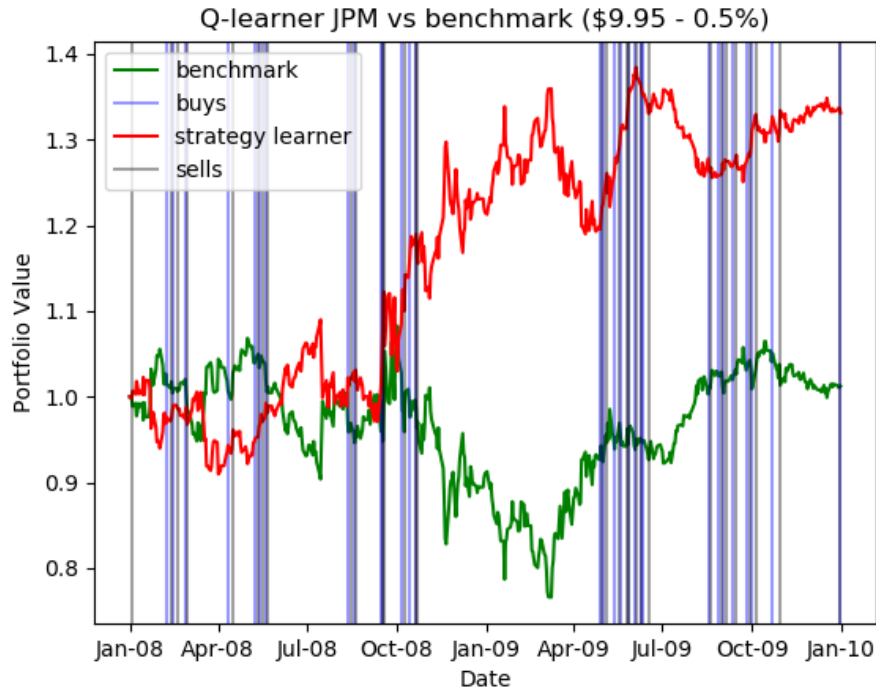


Figure 3

Figure 3 represents the results of the execution of the Q-learner trading strategy trading JPM during the in sample period between January 2008 through December 2009. Starting with an initial cash outlay of \$100,000, the Q-learner strategy had an ending value of \$133,083, while the benchmark had an ending value of \$101,230.

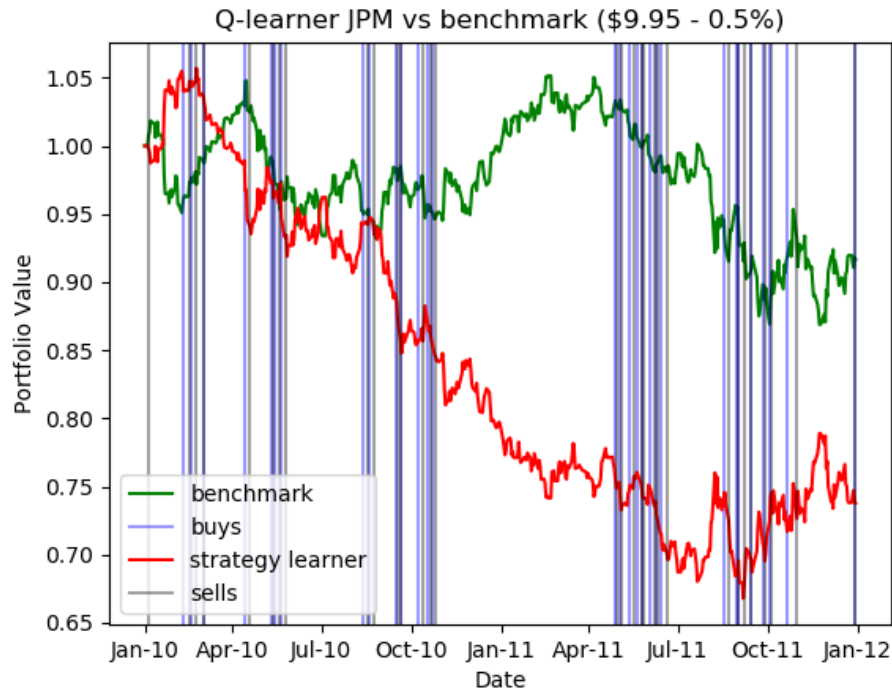


Figure 4

Figure 4 represents the results of the execution of the Q-learner trading strategy trading JPM during the out of sample period between January 2010 through December 2011. Starting with an initial cash outlay of \$100,000, the Q-learner strategy had an ending value of \$73,780, while the benchmark had an ending value of \$91,660.

Below is a table of summary metrics of the Q-learner Strategy vs the benchmark for both the in-sample and out-of-sample periods, including commissions and slippage (impact).

| Metric | Q-learn (in) | Bench (in) | Q-learn (out) | Bench (out) |
|----------------------|--------------|------------|---------------|-------------|
| Cumulative Return | 0.3308 | 0.0123 | -0.2622 | -0.0834 |
| Mean Daily Returns | 0.0006 | 0.0002 | -0.0005 | -0.0001 |
| STD of Daily Returns | 0.0137 | 0.0141 | 0.0101 | 0.0083 |
| Sharpe Ratio | 0.7449 | 0.1542 | -0.8433 | -0.2526 |
| Ending Value | \$133,083 | \$101,230 | \$73,780 | \$91,660 |

Table 2—Manual Strategy vs Benchmark Metrics

5 EXPERIMENT 1

This experiment uses the parameter settings explained in detail in the prior sections. The in sample period is January 2008 through December 2009. Initial cash is \$100,000, and commissions are \$9.95 per trade, with slippage set to 0.5%. A comparison of performance of the Manual Strategy, Q-learner, and Benchmark follows.

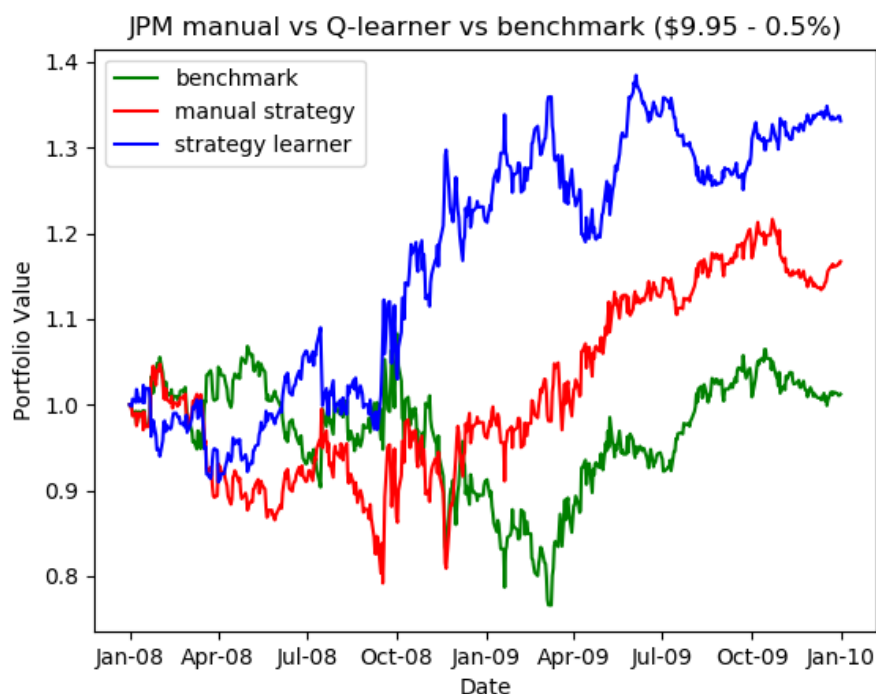


Figure 5

Figure 5 represents the results of the execution of the benchmark, manual, and Q-learner trading strategy trading JPM during the in sample period between January 2008 through December 2009. Starting with an initial cash outlay of \$100,000, the benchmark had an ending value of \$101,230, the manual strategy had an ending value of \$116,703, while the Q-learner strategy had an ending value of \$133,082.

Since a random seed was set in order to obtain the results for the Q-learner, the results will be reproducible based upon that seed. The convergence function is defined as the cumulative returns being within 0.1% of each other for more than 3 consecutive training epochs. It may converge to a local rather than global

minimum, depending on the seed that is set. The author experimented with different seeds and this does appear to be the case. Therefore, the results would not be reproducible over the in sample period if the random seed was not set. The assignment of bins was also done via trial and error to meet the performance requirements of the project and is also arbitrary. The Q-learner is overfit to the sample period because the whole sample period was used to ensure all states during the sample period were present. Dyna was not used in order to meet timeout constraints, which may have resulted in more consistent convergence over different seeds. A better approach may be to limit the Q-learner to just a portion of the in sample period and use data augmentation to fill in the missing states. The Q-learner is also overfit to each symbol's price history. It would be better to train the Q-learner over multiple symbols so it can generalize better over the out of sample periods.

6 EXPERIMENT 2

This experiment tests the hypothesis that commissions and slippage have a negative impact on trading strategy performance on a risk adjusted basis.

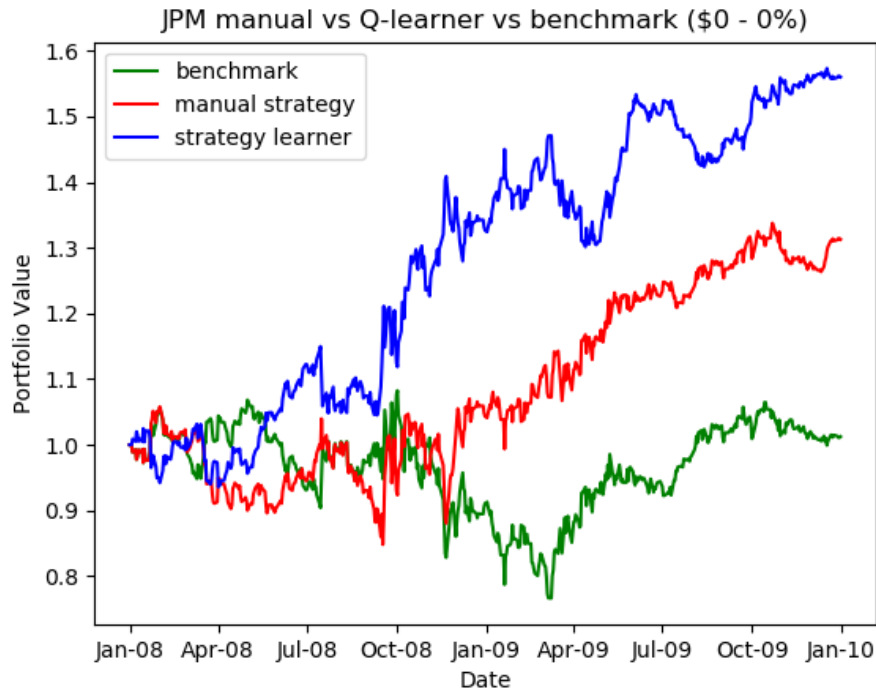


Figure 6

Figure 6 represents the results of the execution of the benchmark, manual, and Q-learner trading strategy trading JPM during the in sample period between January 2008 through December 2009. The same parameters are used as in experiment 1, except it assumes no commissions or slippage. Starting with an initial cash outlay of \$100,000, the benchmark had an ending value of \$101,230, the manual strategy had an ending value of \$131,270, while the Q-learner strategy had an ending value of \$156,000.

The table below contains the summary metrics of executing both Manual and Q-learner strategies without commissions or slippage included.

| Metric | Benchmark | Manual | Q-learner |
|------------------|-----------|-----------|-----------|
| Cum Return | 0.0123 | 0.3127 | 0.5600 |
| Mean Daily Ret | 0.0002 | 0.0006 | 0.0009 |
| STD of Daily Ret | 0.0141 | 0.0156 | 0.0129 |
| Sharpe Ratio | 0.1542 | 0.6545 | 1.1482 |
| Ending Value | \$101,230 | \$131,270 | \$156,000 |

Table 3—Benchmark Metrics vs Manual Strategy vs Q-learner
(no commissions or slippage)

In comparing the summary metrics from Table 3 to the results from the in sample period for tables 1 and 2, the hypothesis is supported.

The cumulative returns for the manual strategy increase from \$116,703 to \$131,270. The mean daily returns increases from 0.0004 to 0.0006. The standard deviation of daily returns decreases from 0.0165 to 0.0156. Meanwhile, the sharpe ratio increases from 0.4148 to 0.6545.

The cumulative returns for the Q-learner increase from \$133,082 to \$156,000. The mean daily returns increases from 0.0006 to 0.0009. The standard deviation of daily returns decreases from 0.0137 to 0.0129. Meanwhile, the sharpe ratio increases from 0.7449 to 1.1482.

Accounting for commissions and slippage decreases overall strategy profitability, while increasing risk.