Report

Yuanzheng Zhu (yzhu319)
yzhu319@gatech.edu

For all tests, Data range: 2008-01-02 00:00:00 to 2009-12-31 00:00:00. Symbol: "JPM". In the chart, green line for "LONG", red line for "SHORT".

1. Experiment 1

1.1. Manual Strategy

Manual Rule-Based Traders (3 strategies based on 3 different indicators: simple moving average SMA, exponential moving average EMA and Bollinger Band value)

SMA: simple moving averages
For instance, if we use 21 days moving average
SMA[today] is the mean of last 21 days' price

SMA_indicator = price/ SMA -1
For instance, choose threshold = 0.1
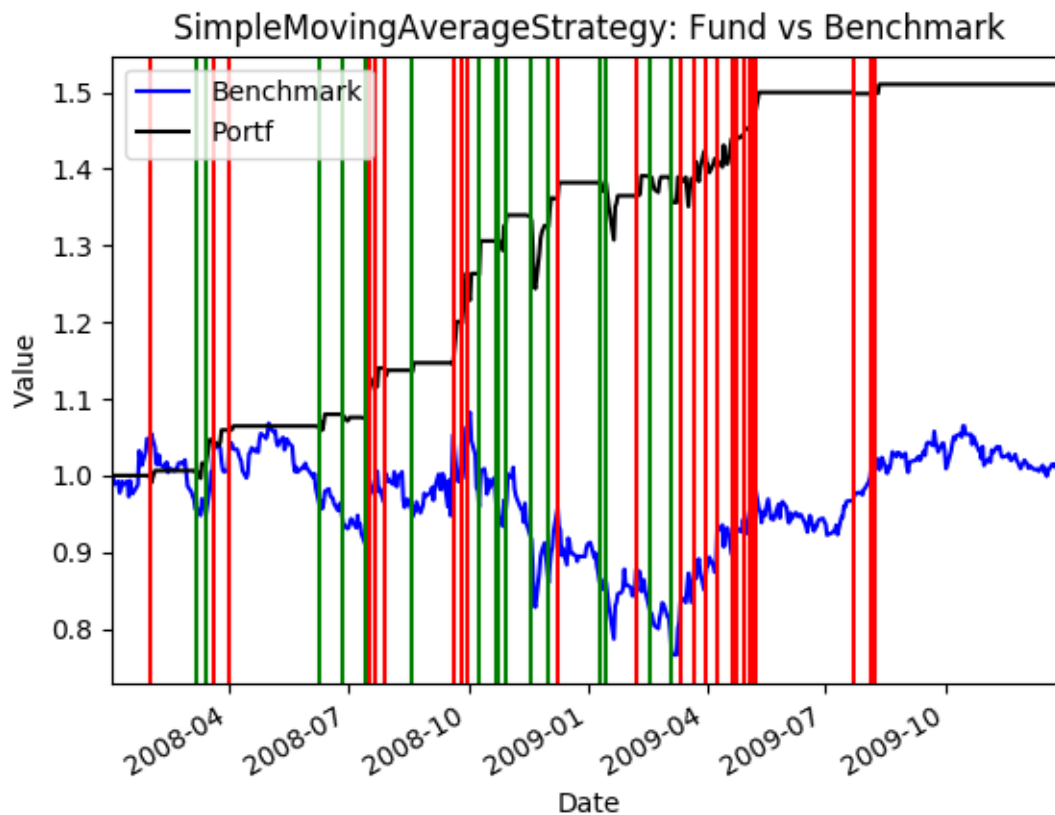If SMA_indicator > 0.1, sell signal
If SMA_indicator < -0.1, buy signal
Otherwise, no signal

SMA results
In-sample data: after tests, best cumulative return: window size N = 21 days, threshold = 0.1

Cumulative Return of Fund: 0.5103215
Cumulative Return of Benchmark: 0.0123

SimpleMovingAverageStrategy: Fund vs Benchmark

EMA: Exponential moving average

Exponential moving average (EMA) reduce the lag by applying more weight to recent prices.
EMA [today] = (Price [today] x K) + (EMA [yesterday] x (1 – K))
Where:
- K is multiplier, K = 2.0/(N+1)
- N = Lentgh of EMA, usually take 10~20

The very first EMA value is calculated by the average of the first N days' data.

EMA_indicator = price/ EMA -1
For instance, choose threshold = 0.08
If EMA_indicator > 0.08, sell signal
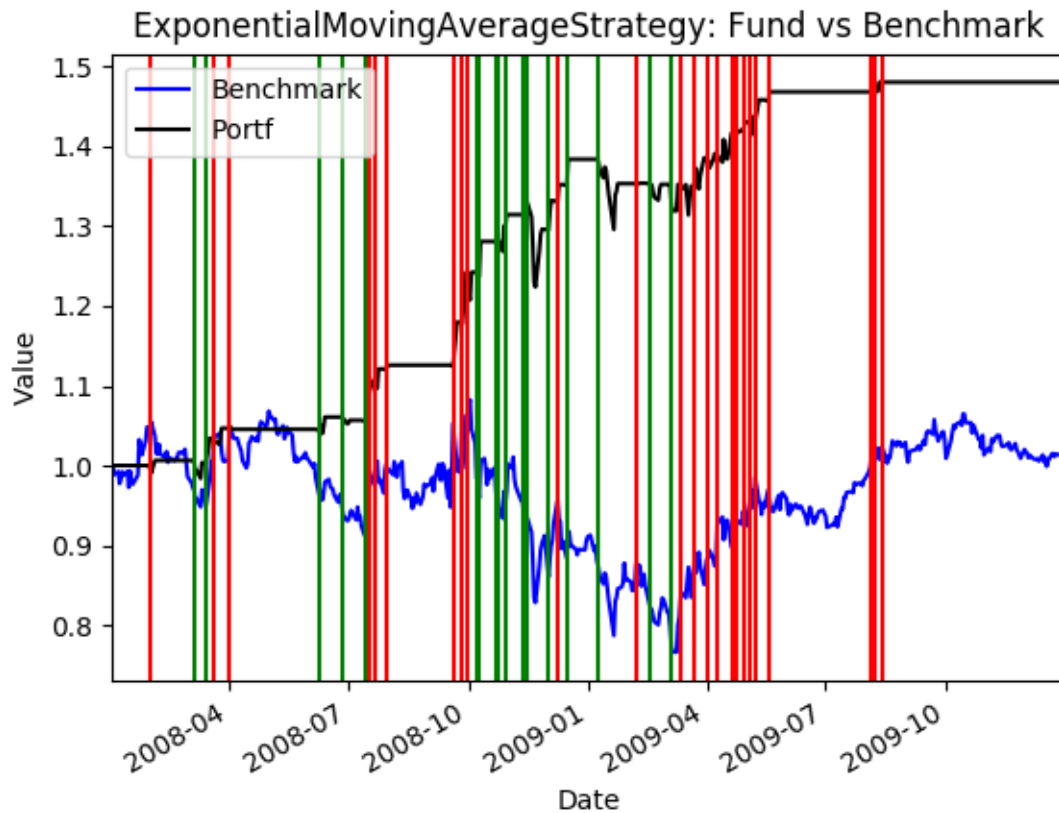If EMA_indicator < -0.08, buy signal
Otherwise, no signal

EMA results

In-sample data: after tests, best cumulative return: window size N = 20 days, threshold = 0.08

Cumulative Return of Fund: 0.480659

Cumulative Return of Benchmark : 0.0123

Bollinger Bands
For instance, choose parameters: 20 days window size, threshold bandwidth is 1.2 times standard deviation.
Based on simple moving average, the band is 1.2*sigma (rolling standard deviations of past 20 days) above and below rolling mean curve, we obtain upper band and lower band.

Bolinger_Bands_Index = (price – moving_average) / (2*moving_std)

If Bolinger_Bands_Index > 1, sell signal
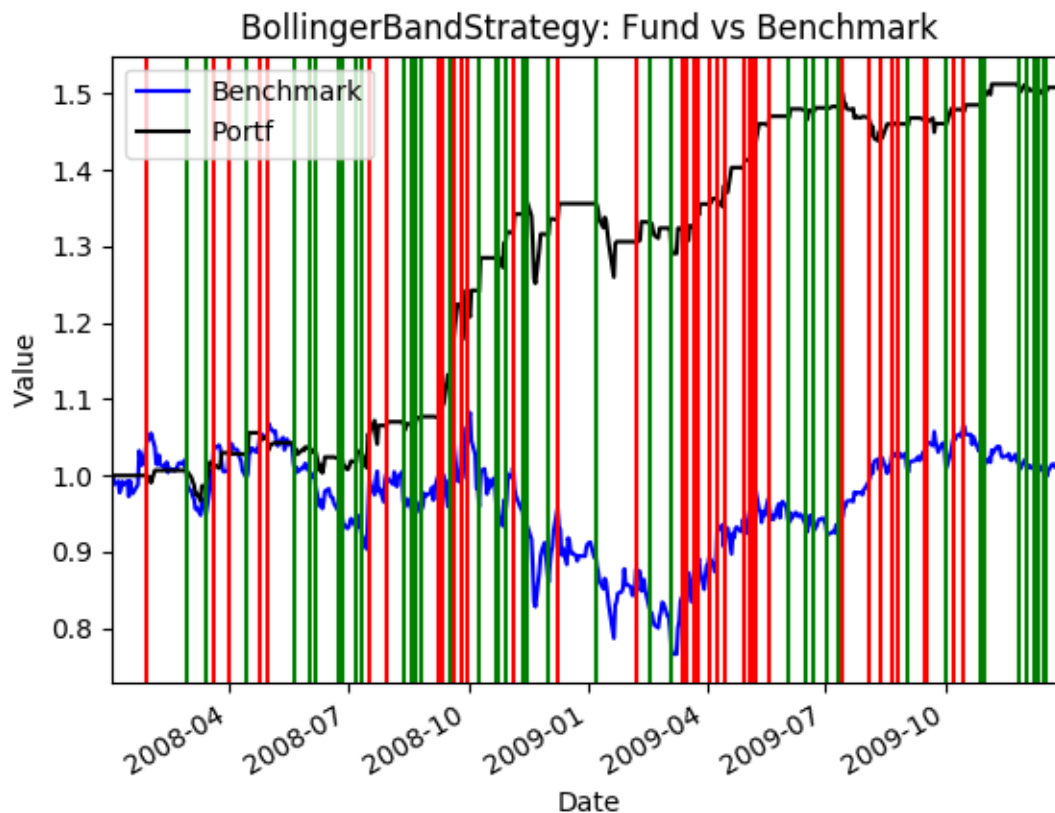If Bolinger_Bands_Index < -1, buy signal
Otherwise, no signal

Bollinger Band results
In-sample data: after tests, best cumulative return: window size N = 20 days, band width = 1.2 (1.2 times rolling_std)
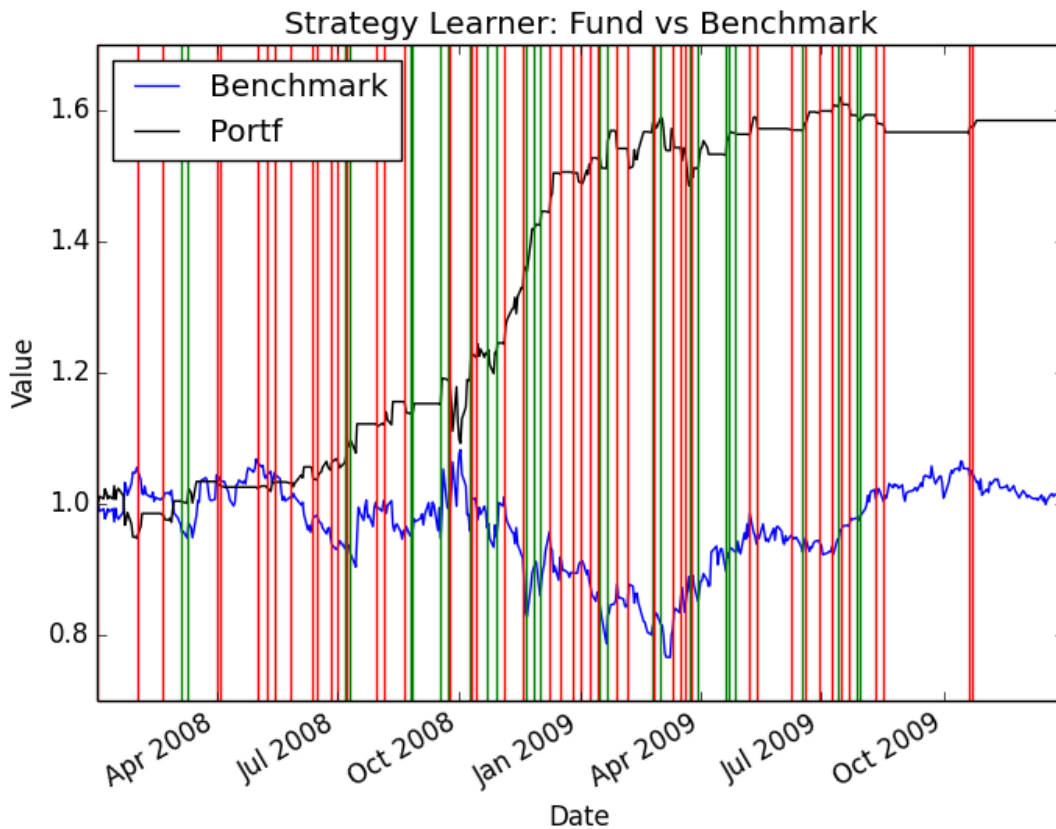
Cumulative Return of Fund: 0.5083125
Cumulative Return of Benchmark: 0.0123

BollingerBandStrategy: Fund vs Benchmark

1.2. Machine Learning based strategy

Use the same indicator parameters as mentioned before (like window size, band width) in the manual strategy to calculate indicator values for each day. X data for each sample (day) are the values of these 3 indicators. The Y data (or classifications) is based on N day return (**N= 8**). Y value is classified as a +1 or "LONG" if the N day return exceeds a threshold YBUY (**YBUY= +0.05**). Also, classify Y as a -1 or "SHORT" if the N day return is below a threshold YSELL (**YSELL=- 0.05**). Otherwise Y is classified as a 0 or "CASH".
Use the data to train a Random Forest Learner with **bag =20** and **leaf size = 6**. Then use this learner to make Y predictions with a new set of (X1, X2, X3).

Cumulative Return of Fund: 0.584512309427
Cumulative Return of Benchmark: 0.0123

1.3. Comparison and discussions

	Strategies	Cumulative return
Manual Strategy	SMA	0.5103215
	EMA	0.480659
	Bollinger Band	0.5083125
Machine Learning	Random Forest Classifier	**0.5845123**
	Benchmark	0.0123

1.3.1. Describe your experiment in detail: Assumptions, parameter values and so on.
- See detailed description above.
1.3.2. Describe the outcome of your experiment.
- Comparing the cumulative returns, the classification learner based on 3 indicators is consistently performing better than the manual strategies based on a certain indicator.
1.3.3. Would you expect this relative result every time with in-sample data? Explain why or why not.

- I expect this relative result every time with in-sample data, since the random forest classifier takes advantage of all three indicators, thus less biased towards one indicator.

2. Experiment 2

For this experiment only, set a fixed value to seed =123. Commission = 0.
The parameter "impact" will essentially affect the YBUY and YSELL, i.e. the threshold to trigger a "LONG" or "SHORT". If market impact is high, threshold is (YBUY+ 2*impact) is higher to trigger a "LONG" (similar for "SHORT"). When we keep other parameters the same (bag=20, leaf=6, N-days return N=8, YBUY=+0.05, YSELL=-0.05), and test impact values = (0.000, 0.001, …, 0.009, 0.010).

Impact value (X)	Cumulative Return (Y1)	Number of trades (Y2)
0.000	1.044	181
0.001	0.784	207
0.002	0.605	189
0.003	0.603	188
0.004	0.772	171
0.005	0.563	163
0.006	0.3523	149
0.007	0.3297	143
0.008	0.3242	139
0.009	0.2964	140
0.010	0.2598	140

As market impact (X) increases, the threshold for triggering a trade is higher, thus less number of total trades (Y2). So, the average duration between two trades will be longer. For in-sample data, this means we sacrifice some of the near future returns in the hope of getting higher returns by waiting longer. This will result in less cumulative return for in-sample data, but may be better for out-of-sample data. In the extreme case, impact value X=0 or N = 1 will give us the best in-sample results, as we get the most of return for each day, like the "best possible strategy", but this model will perform poorly in out-of-sample data as it over-fits the training data.