Short Term Trading Models Using Hurst Exponent and Machine Learning

Gursewak Singh Sidhu(sidhus234@gmail.com),

Ali Ibrahim Ali Metwaly(aiam_2000@hotmail.com)

Animesh Tiwari(animesh_t7@hotmail.com)

Ritabrata Bhattcharyya(ritabrata.bhattacharyya@wqu.org)

Abstract

Predicting the direction of Stock Indices has always been an appealing topic which has motivated researchers over the years to develop better predictive models. Recently, Machine learning (ML) based models have been frequently deployed to forecast the direction of classic financial time series data. In the 1950s, Hurst Exponent was introduced as a statistical measure to classify various Time Series. This research analyzes the effectiveness of using Machine Learning and Hurst Exponent along with popular Technical Indicators for short term trading predictions. In this study we explore the use of Hurst Exponent to segment data for a short-term machine learning model in order to improve trading strategy. A comparative analysis has been carried out between the performance of a standalone short-term model, and a Segmented model (Segments based on hurst exponent cut off) in S&P 500, SSE Composite Indices, Gold SPDR Shares and Bitcoin. This new approach is being introduced in order to reach the optimum integration between Machine learning & Hurst Exponent.

Keywords: *Short term trading models, Hurst Exponent, Technical Analysis, S&P500, Machine learning for finance, Financial time series, buy-and-hold, stop-loss, cryptocurrencies, Bitcoin, BTC-USD*

1. Introduction

The project explores usage of Machine Learning models to perform classification based supervised learning where the outcome of interest is to predict the direction of asset's movement from short-term perspective. For this purpose, last 10 years of data, essentially covering the post Global Financial Crisis era is considered. The research focuses on analyzing the effectiveness of Machine Learning in trading by breaking the problem into two smaller steps, one to understand the medium-term trend (using Hurst Exponent), and a machine learning model to predict the short-term action (Buy, Sell, None).

This study relies on two pillars, the first pillar utilizes Hurst Exponent to understand the medium-term trend. The second pillar utilizes Machine Learning, (Technical Indicator based Random Forest, Gradient Boosting Machine and Xgboost Classifier models) to predict the short-term trend. The amalgamation between Hurst Exponent output-based segmentation and Machine Learning is the first approach.

2. Theoretical Framework

2.1 Literature Review

In 1951, British hydrologist, Harold Edwin Hurst introduced Hurst Exponent which quantifies the long-term memory of a time series. Varying between 0 to 1 it explores the autocorrelations and lags in a Time Series. Hurst proposed that H = 0.5 represents no correlation which implies no association between the current value and the past values of the series. For H between 0 to 0.5, the series behaves in a mean reverting fashion alternating between highs and lows. As hurst exponent approaches zero, the series mean reverting strength increases. When H is between 0.5 to 1 it implies long term, positive correlation displaying a rather consistent movement in upward or downward direction. Such series show trend for some time, and then may see a certain abrupt disruption. As value approaches one, the series trend behavior becomes dominant. The Hurst Exponent method from hydrology was later introduced and explored in other fields by Mandelbrot & Ness [2], Mandelbrot [3] and others. Peters [4] introduced Fractal Markets Hypothesis (FMH) which affirms non-normality and autocorrelation of asset returns. FMH allows for a broader range of returns' behavior. Some other papers which explored the use of Hurst Exponent for financial series are:

Lipka and Los (2002) [5] explored Hurst Exponent's ability to measure long term dependence in data series for eight European stock market Indices. Lipka and Los observed that the FTSE100 returns represent highly-efficient market with rapid mean reversion compared to a Geometric Brownian Motion.

Corazza and Malliaris (2002) [6] explored the application of Hurst Exponent in the forex markets. Corazza and Malliaris found Hurst Exponent to be significantly different from 0.5 in many of the samples. The authors also observed that Hurst Exponent value for foreign currency markets change over time, and is not fixed. This could indicate that foreign exchange markets are inefficient, compared to their equity peers.

Cajueiro and Tabak (2004) [7] used a rolling window approach in order to calculate Hurst Exponent to assess the efficiency of 11 emerging markets and compare the same with US and Japan (which are assumed to be efficient markets). They found that Asian equity markets show less efficiency than Latin American markets, whereas the developed markets are more efficient (Hurst Exponent = 0.5).

Mitra [8] explored the values of Hurst Exponent over varying window sizes. They analyzed the value of Hurst Exponent for twelve stock index series for past 10 years. The author showed that in long window sizes the Hurst Exponent value is very close to 0.5, but over small window sizes of 60 days, its value varies widely. Mitra observed the correlation between H-value and returns, and hence H-value can be used in tandem with Technical Analysis. The authors summarized that though H-value in itself might not predict the direction of a trend, but when H value is high then trend detecting rules could provide higher returns.

In [9] authors showed that irregularities in the series, such as jumps or spikes lead to inappropriate scaling and hence an incorrect estimation of the Hurst Exponent. The authors also pointed out the statistical error while measuring Hurst Exponent in the presence of self-affine fractal scaling.

Kirichenko, Radivilova, and Deineko [10] described different methods that are the commonly used for estimating the Hurst exponents (R/S-analysis, detrended fluctuation analysis (DFA), variance-time analysis, and wavelet-based estimation). The results showed that the estimates of the H exponent are biased which depend on the true value of the degree of self-similarity of the process as well as the length of Time Series.

Morales, Matteo, Gramatica, and Aste [11] investigated the use of the Hurst Exponent, computed over a cascading time-window, to evaluate the level of stability of financial firms. Financial firms bailed-out post the 2007-2010 credit crisis show a clear increment in generalized Hurst Exponent in the period prior to crisis.
Existing research shows that estimation of Hurst Exponent over smaller time frames is volatile, and needs to be handled with care. Though on a long-term basis the markets are efficient and have Hurst Exponent of 0.5, the smaller term tells a different story.

It is not new for researchers to find the stocks and stock market indices appealing from prediction perspective. While it all begins as a researcher's quest to solve this problem, there are many involved parties in the market that get affected by the outcomes, and hence the ability to foresee potential market movements in the short to medium term horizon could present lucrative trade opportunities. There are two broad approaches:
First approach assumes that the index has everything baked-in i.e. the index at time t represented by Y_t is a composite of various underlying factors, hence, we only consider this value to be meaningful while predicting a future value after k time intervals Y_{t+k}. This is time series analysis-based approach.
Second approach attempts to explore the relationship of the form $Y = f(x)$ where the index Y is seen as the Effect of certain causal factors or Xs. Here we attempt to understand the explanatory variables which in turn help us predict the response Y. This is Machine Learning based approach where we try to create various models and choose the ones which serve high and sustainable accuracy while predicting the historical as well as future outcomes of interest. This is the approach we intend to explore in our project. We explored the existing work done by various researchers in this area.

In the paper [12] published by Pyo S 2017 various technical indicators were modelled using non-parametric machine learning models like ANN, and SVM with polynomial as well as RBF kernel. In addition, interesting hypotheses were formulated and validated e.g. taking google trends as one of independent variables, analysis of the top contributing listed companies on the overall index, and a rollover strategy while considering the train and test sets over a rolling window made the results more reliable.
This paper takes [13] as a reference and calls out the experimental procedures used in [13] impractical to investors because while deciding the training and test data time series nature of the data was completely discarded.

Another interesting research in this area [14] by Qui M 2016 for the NIKKEI 225 focuses on creating same model on two non-overlapping input sets, wherein the optimization of single hidden layer ANN using Genetic Algorithm substantially improves the predictive power measure in terms of Hit Ratio. Thus, the right features and a tuned model seems to work. However, there are many uncertainties and apprehensions, these approaches have not been applied on the index we are trying to predict.

Also, markets evolve over a period of time, for example, the GFC had an impact which lasted a few years, whereas during the pandemic which is the worst ever crisis we are witnessing in our lifetime, the markets took just a few months to bounce back. Therefore, the stage at which the predictions are being done and the markets considered in scope could also have a bearing on the performance of quantitative algorithms.

Martinez, Lisana B [15] 2018 considered the impact of financial crisis on the 6 EU countries, found that stocks seem to respond to the crisis built-up in a smoother way compared to the bonds this is validated using the Hurst Exponent. But we can't say the same holds good for the pandemic because the time to react in this case was rather limited. The same is found in the research carried out by [16] Matthieu Garcin 2020.

In [17] authors propose a modern deep learning approach integrating wavelet transforms (WT) along with stacked auto-encoders (SAEs), and long-short term memory (LSTM) for stock price forecasting. The deep learning framework comprises three stages. In first stage the stock price information decomposed using wavelet transform to remove noise. In second stage auto-encoders are used to generate feature set while minimizing root mean square error between input and output. In final stage of the framework, the authors used an LTSM model to predict the subsequent day's closing price. The performance of the model was compared for six different market indices and the results showed that the model outperformed the similar approaches in both profitability and predictive accuracy.

In [18] study focuses on use of deep learning to predict future stock prices. The authors designed a deep learning model with 715 inputs based on technical analysis. They also filtered the data pattern in price fluctuations to get similar data input to models. Different combinations of input features and target vectors were explored to understand the performance of deep learning for stock price prediction.
This [20] paper presented a comparative study for ARIMA, Neural Network, and stochastic geometric Brownian motion for stock price. Their research concludes that compared to the neural network model, the statistical model, and the stochastic model perform better at predicting next day stock price.

In [21] the authors explore the usage of feature selection for data processing using two neural networks for predicting returns, and a simple rule-based system to integrate the two on S&P500. The purpose of this study was to explore if more systematic data filtering and return integration can bring improvements. Multiple combinations were explored and validated for 5-year trading period.

2.2 Competitor Analysis

There are multiple research papers on use of Hurst Exponent for trading, and to identify market cycles. Existing research shows that Hurst Exponent is widely used in the time series data pertaining to physical systems, healthcare and finance. Use of machine learning in Short term trading is a well-documented field. Wang [22] used PCA-SVM integrated model to predict the stock market indices. Both internal and external financial factors were explored in the model. One major drawback of the given approach is using 1 for price increase, and -1 for price decrease. The third state of oscillations is not being considered. In [23] Qian, & Rasheed explore the use of Hurst Exponent as an indicator for short term trading. Though the exponent may be a good indicator but given today's common use of quant algorithms, a simple rule-based system might not offer an edge in the market. Chenoweth, and Obradovic in [21] built a two layered structured model where the incoming data is divided into market trends, and then a prediction model using neural networks. The major shortcoming of this system is that the medium-term trend indicator is based on a simple rule, and one major drawback of ANNs is that the performance on unseen samples declines rapidly when the ANN model is over-fitted to training observations e.g. the noise in stock information could encourage ANNs to create a rather complex model which overfits.
Our novel approach explores the use of Hurst Exponent as a market trend Indicator. It has been further observed that while Hurst Exponent in itself might not directly predict the direction, in combination with technical analysis, can show better efficiency in the short term when H is approximately equal to 0.5. Also, in small window sizes (which is more relevant for Short term trading, and medium-term trend), Hurst Exponent estimation is

biased. Therefore, using technical indicators as predictors in short term model seems logical. The study also explores if the window size affects the accuracy of medium-term trend.

During literature review we didn't come across any one stop solution that utilized combination of Hurst Exponent and Technical Indicators using two separate models for medium and short term. Together with this we are breaking the problem into simple steps where we are benchmarking Machine Learning classification models on short term between one single model versus Hurst Exponent based segmented models.

In most research we came across, the researchers use the absolute value to define uptrend or down-trend. We shall explore a more dynamic way of defining a trend in short-term using rolling average standard deviation; rather than a fixed approach of either a certain percentage increase/decrease or just simple increase or decrease. Our proposed system also accounts for an oscillation phase in markets. The same period is skipped entirely with no traditional activity.

3. Methodology

3.1 Data

The data sample is focusing on last 10 years of data of S&P500 index, SSE Composite index, Gold (SPDR Gold Shares), and Bitcoin USD (BTC-USD). Data was gathered from Yahoo finance. SSE Composite Index is the most commonly used indicator to reflect SSE's market performance. Constituents for the SSE Composite Index are all listed stocks (A shares and B shares) at the Shanghai Stock Exchange. S&P500 represents Free-float capitalization-weighted index in US Market. SPDR Gold shares seek to reflect the performance of the price of gold bullion, less the expenses of the Trust operations. The Trust holds gold bars and from time to time, issues Baskets in exchange for deposits of gold and distributes gold in connection with redemptions of Baskets. Bitcoin (BTC) is a cryptocurrency which uses peer-to-peer technology to operate with no central authority or banks; managing transactions and the issuing of bitcoins is carried out collectively by the network.

The whole sample was split into two sets, training set, and test set. The structure for model validation, and testing is defined below. "Cross Validation" technique implemented in scikit-learn package was used for model selection in Grid Search. The quant strategy was tested with out of time sample as below:

Train data (2010 - 2018)	Test data (2019 - 2020)

Figure 1. Train and Test samples

3.2 Segmentation based on Hurst Exponent

Hurst Exponent is calculated via rolling window with varying sizes (100, 200, 300 and 400 trading days). Rescaled range hurst exponent for series $X_1, X_2, \ldots, X_n$ is calculated as below:

1. Calculate the mean of series

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. Create a mean adjusted series (Y) as

$$Y_t = X_t - m \quad for\ t = 1,2,\dots,n$$

3. Calculate cumulative deviate series Z

$$Z_t = \sum_{i=1}^{t} Y_i \quad for\ t = 1,2,\dots,n$$

4. Create a range series R:

$$R_t = \max(Z_1, Z_2, \dots, Z_t) - \min(Z_1, Z_2, \dots, Z_t) \quad for\ t = 1,2,..n$$

5. Create a standard deviation series R:

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(X_i - u)^2} \qquad for\ t = 1,2,\dots n$$

Where:

$$h\ is\ mean\ for\ time\ series\ values\ X_1, X_2, \dots, X_n$$

6. Calculate the rescaled range analysis (R/S):

$$(R/S)_t = \frac{R_t}{S_t} \quad for\ t = 1,2,\dots,n$$

7. The Hurst Exponent is estimated by fitting the power-law $E[R(n)/S(n)] = C \times n^H$ to the data. This is done by taking the logarithm of both sides, and fitting a straight line. The slope of the line gives H
8. Anis-Lloyd corrected R/S Hurst Exponent is calculated as 0.5 plus the slope of $\frac{R(n)}{S(n)} - E\left[\frac{R(n)}{S(n)}\right]$

Visual graph for the whole time period has been used to select the optimal window size. After selecting optimal window size, two new variables were defined; 1. Ratio of Number of days with increase in value over last day to days with decrease in value over last day (Gives an indication of trend behaviour of time series), 2. Ratio of number of days when the price went from increasing to decreasing to number of days it went from decreasing to increasing (gives an indication of mean reversion behaviour of time series). Plotting above variables for different values of Hurst Exponent, a cut-off was estimated to divide the data into two segments.

3.3 Model Selection Roadmap

Most research in stock trading using quant models focuses on single stage models to predict next day closing prices. Our unique approach is to break down the problem in two components, one using Hurst Exponent to identify trending behaviour and then over lay that with a specific model to define the trading strategy over short term.

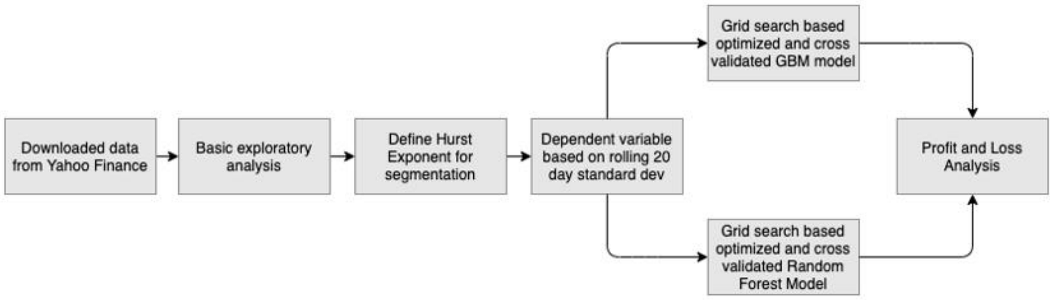

Figure 2. Overall process flow for the trading strategy

The performance for the strategies has been compared for S&P 500, which being a developed and mature market is considered efficient, SSE Composite (China) which is a developing market and is semi-efficient, Gold as a representative of a mature commodities market and Bitcoin, which is emerging cryptocurrency market in early stages (mature compared to other cryptocurrencies available in market).

1. **Option 1:** Use of Hurst Exponent (H) to establish a medium-term behavior, build and train specific models for mean reverting, and trending component for short term trading. Using Hurst Exponent as an Indicator for market in medium term memory of the series.

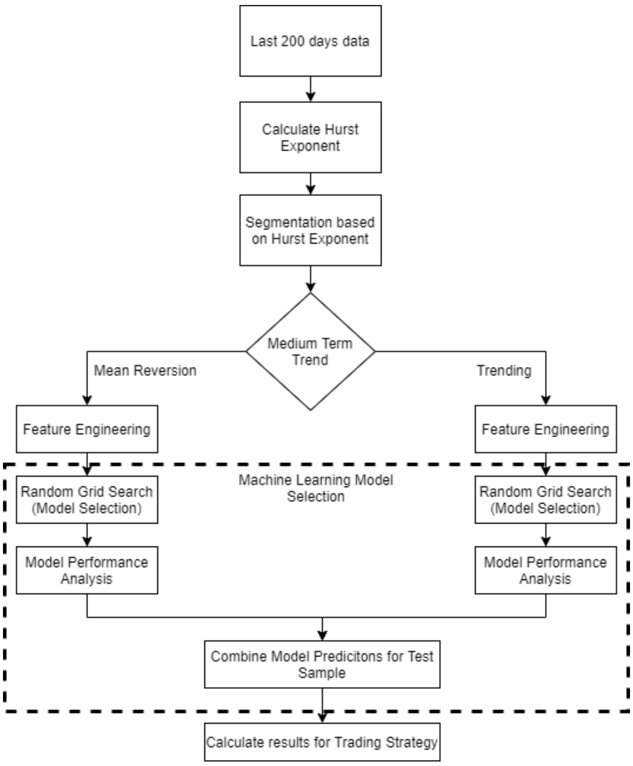

Figure 3. Hurst Exponent based Segmented Model approach

2. **Option 2:** In this option the whole dataset was used to build a short term trading model.

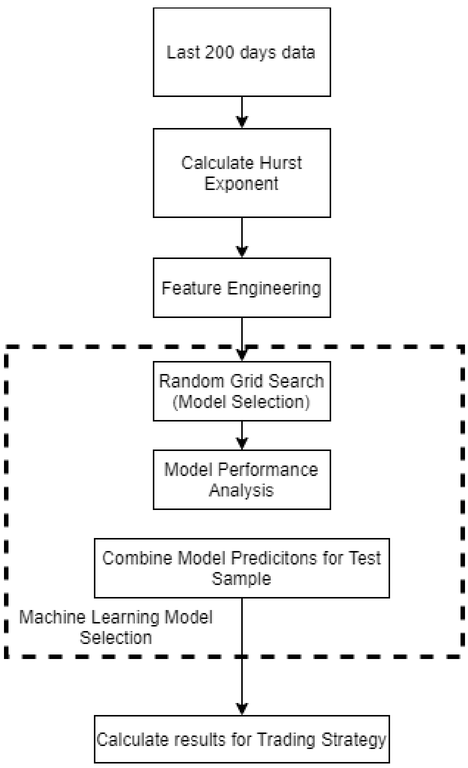

Figure 4. Machine Learning Model (without segment)

The above two structures differ basis the use of medium-term trading model. The first one explores the use of Hurst Exponent to identify the medium-term trend, whereas in the second, a short-term machine learning model has been built on whole dataset.

The classification objective for short term model was defined in below manner:

Step 1: Calculate last 20 trading days rolling window standard deviation

Step 2: Calculate upper and lower bound for next 5 days as:

- Upper bound = Current day price + 1 standard deviation
- Lower bound = Current day price – 1.5 standard deviation

Step 3: The classification variable is tagged as:

- 1: if the stock price (or Index value) breaches the upper limit in next 5 days
- -1: if the stock price (or Index value) breaches the lower limit in next 5 days
- Else 0

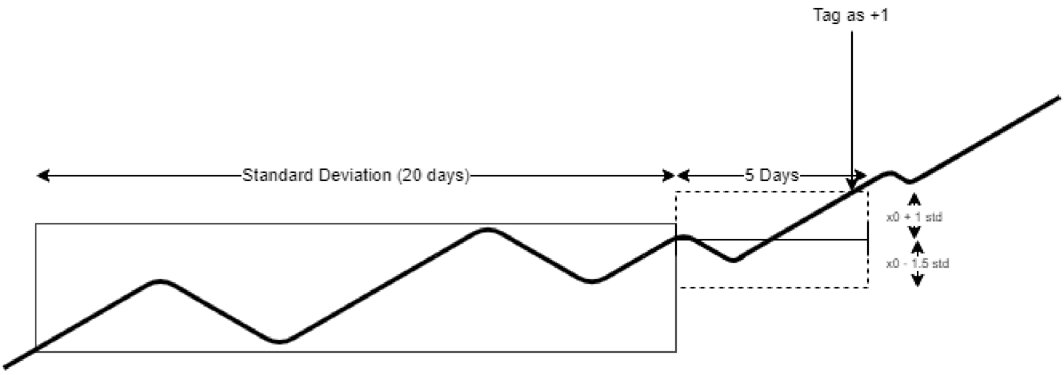

Figure 5. Dependent Variable definition for short term trading model (Buy signal)

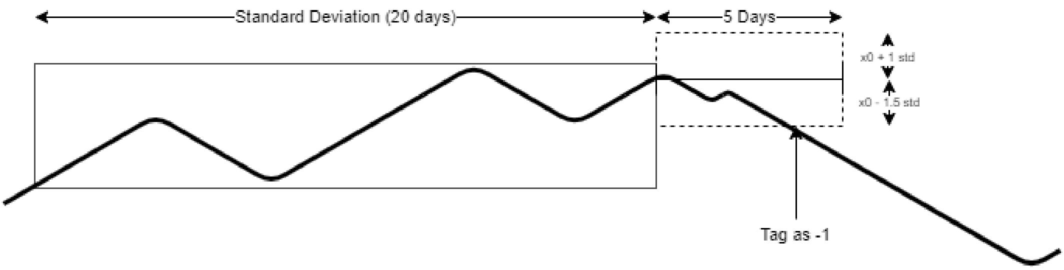

Figure 6. Dependent Variable definition for short term trading model (Sell signal)

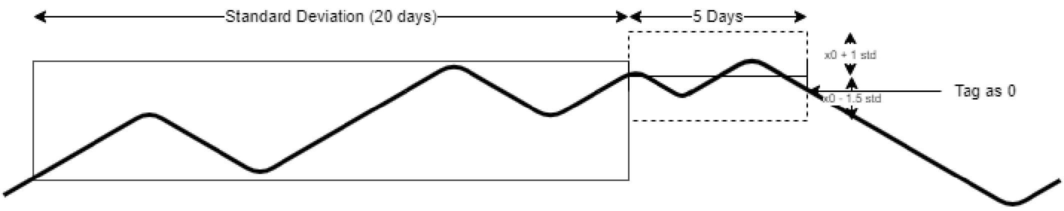

Figure 7. Dependent Variable definition for short term trading model (Discard/No Action)

The above classification problem has been solved using machine learning models Random Forest, and Gradient Boosting Machines Classifiers based on technical indicators.

Upon completing the literature review, we could not find a single research focused combining Hurst Exponent, technical Indicators in medium term model, and short-term Machine Learning segment model to improve trading results. We tried to address this area in our current project. The trading structure results have been compared with one another, and a basic simple strategy of Buy and Hold.

3.4 Trading Strategy

Our trading strategy is based on the outcome of Short-term model, which can give a Buy Signal, a Sell Signal or No Action. A position is initiated based on the model output and is kept open for a maximum of 5 days. If neither the target or stop loss is hit in next 5 days, the position is closed. In case the model gives an opposite direction within those 5 days (before target or stop loss is hit), the position is closed and opposite position is opened. More details can be found below:

1. Buy Signal on Day 0:
    - Target = Buy Price + one standard deviation (of last 20 days)
    - Stop Loss = Buy Price – 1.5 standard deviation (of last 20 days)
    - If the target or stop loss is hit in next 4 days, exit the position
    - If model gives "Sell Signal" in next 4 days, exit the position and take alternate position
    - If in next 4 days no action could be taken because of above, close the position
    - If a new buy signal arises within the 5-day period, reset the days to "1"
2. Sell Signal on Day 0:
    - Target = Buy Price – 1.5 standard deviation (of last 20 days)
    - Stop Loss = Buy Price + one standard deviation (of last 20 days)
    - If the target or stop loss is hit in next 4 days, exit the position
    - If model gives "Buy Signal" in next 4 days, exit the position and take alternate position
    - If in next 4 days no action could be taken because of above, close the position
    - If a new sell signal arises within the 5-day window, reset the days to "1"
3. No Action Signal on Day 0:
    - No position is taken, and existing position is continued

3.5 Environment of the project

For the scope of project, we shall be using Python 3.x version, with below libraries for specific things. Data will be sourced at daily level from yahoo finance (using python api).

library	Usage
Pandas	Data manipulation and testing strategy
Matplotlib & seaborn	Visualizations
Numpy	Numerical calculations (ones which can't be handled in pandas)
SciPy	Scientific functions required for exploration, and analysis
Scikit-Learn	Statistical and Machine Learning model (Training, testing, and performance metrics)
TensorFlow & Keras	Neural network (Autoencoders, and Deep learning classification network)
Hurst	calculation of Hurst Exponent
XGBoost	For xgboost model
ta	Technical Indicators Library

Table 1: Python libraries

4. Results

Data for S&P 500, SSE Composite Indices, SPDR Gold Shares and Bitcoin (BTC-USD) were downloaded from Yahoo Finance from Jan 01, 2011 to Dec 31, 2020. Based on the below plot, Hurst Exponent with window size of 200 was selected for segmentation. Hurst Exponent is very volatile with window size of 100, but from window size 200 onwards the trend remains same, and it grows smoother with higher window sizes. Hence, 200 days trading window was chosen to calculate Hurst Exponent for medium-term trend. Detailed graphs for each asset can be found in [24].

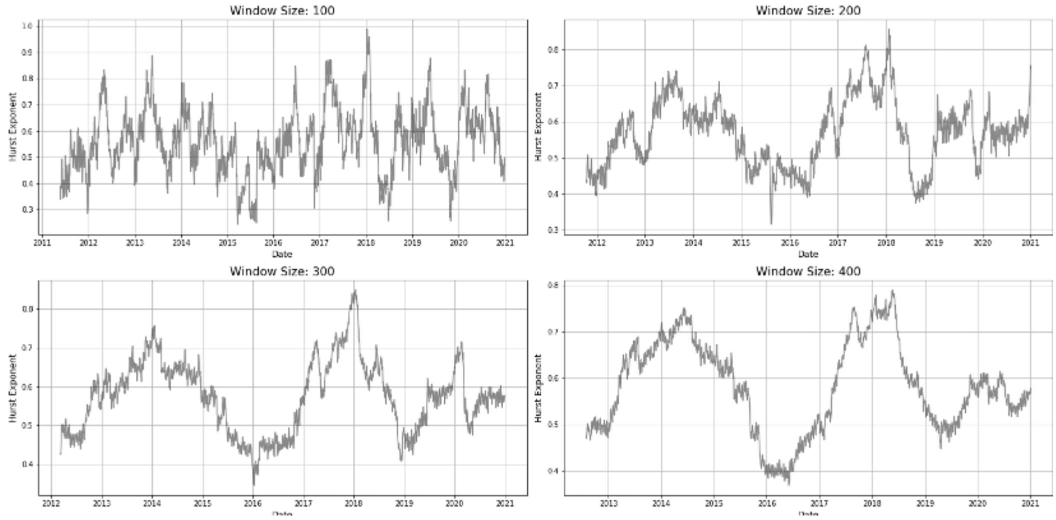

Figure 8. Hurst Exponent for different window sizes for S&P 500

After selecting the optimal window size, Trend ratio variable, and Zero Cross rate variable (as described in section 3.2) were plotted to estimate Hurst Exponent threshold to divide data into segments. For S&P 500, SSE Composite and Gold, based on Hurst Exponent with window size 200 trading days, the cut off was estimated at 0.6, and for BTC-USD 0.735 was defined as threshold. Among the asset classes, Bitcoin behaved most differently. Being a new emerging asset class (and one of the most mature cryptocurrencies, both in terms of market valuation, and history), Bitcoin seems to be most trending asset.

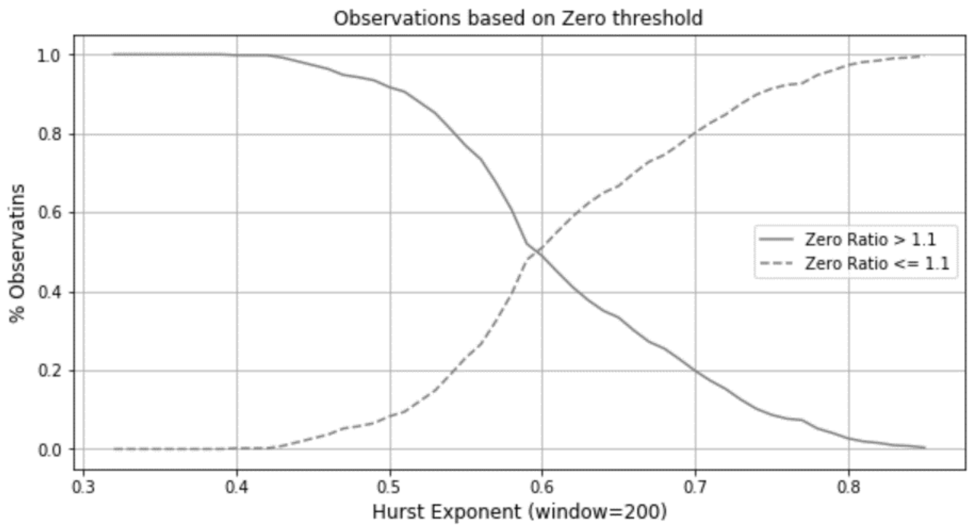

Figure 9. Ratio of Positive to Negative Trend days for S&P 500

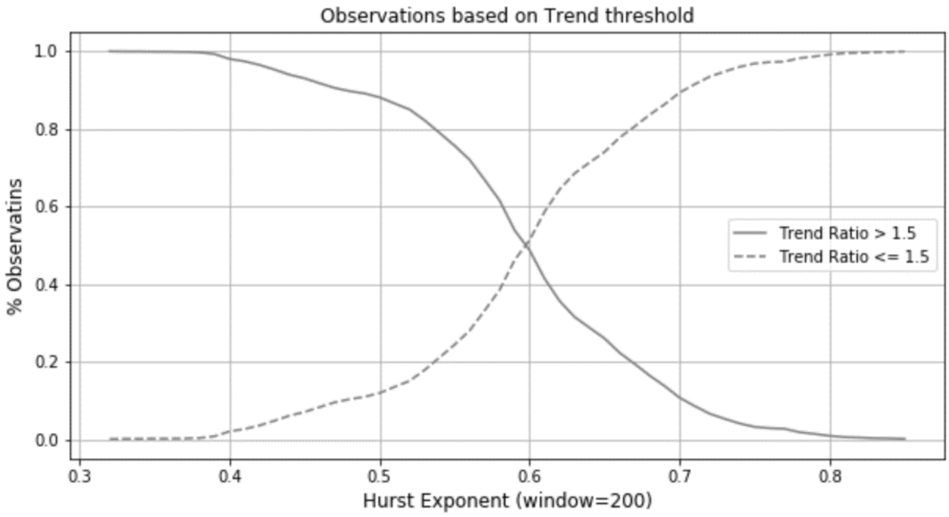

Figure 10. Ratio of zero cross over (Positive to Negative and Negative to Positive) for S&P 500

Based on above two plots (Figure 10, and Figure 11) Hurst Exponent value of 0.6 was decided as cut-off to divide data in two segments. The charts are similar for SSE Composite, Gold SPDR Shares and Bitcoin (BTC-USD) as well, and can be found in[24]. For feature engineering, we used the Python library "ta" which generates Technical Indicators. "ta" library in python implements Momentum, Volume, Volatility, and Trend Indicators. Over these indicators we added some simple variables like Simple Moving Average, Maximum, Minimum, Ratio of Minimum to Maximum and Standard deviation on rolling window analysis. To improve the model performance, use of PCA was also explored in feature engineering. After above steps, we are left with below two structures to build our ML based trading strategy.

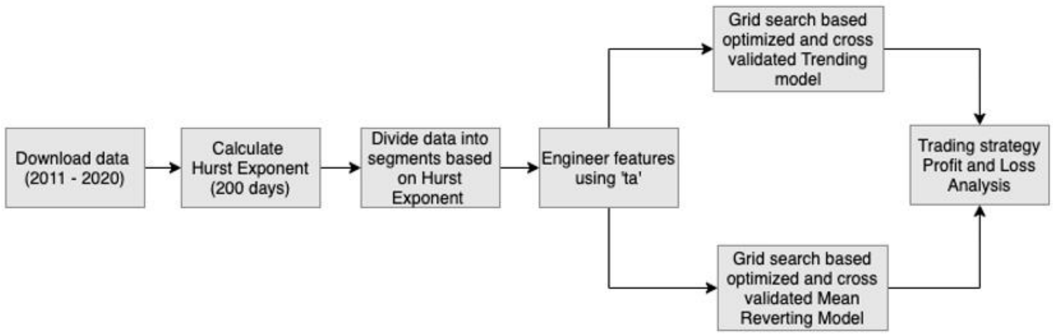

Figure 11: Hurst Exponent Based Segmented Model Approach

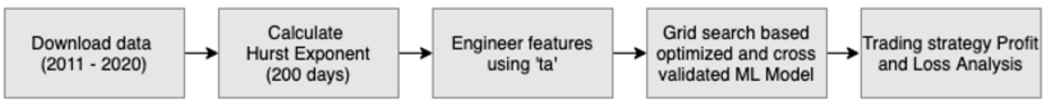

Figure 12: Non-Segmented Model Approach

Scikit-Learn package was used to develop Random forest, Gradient Boosting Models and XgBoost model (Scikit-learn wrapper) for the short term on both the indices. The results for trading strategy in test sample are:

	Random Forest		GBM		XgBoost	
Year	Whole	Hurst segment	Whole	Hurst segment	Whole	Hurst segment
2019	6.13% (6.57%)	-8.90% (6.69%)	-17.07% (9.07%)	-7.87% (9.07%)	-0.83% (6.41%)	5.92% (6.91%)
2020	5.86% (12.04%)	-0.88% (12.65%)	-48.62% (14.36%)	-20.04% (19.24%)	-12.39% (9.35%)	3.14% (12.13%)

Table 2: Results (Annual Returns/Max Drawdown) for S&P 500 Index

	Random Forest		GBM		XgBoost	
Year	Whole	Hurst segment	Whole	Hurst segment	Whole	Hurst segment
2019	2.20% (10.64%)	-4.08% (12.36%)	-3.97% (10.56%)	9.87% (11.51%)	-5.27% (11.14%)	-15.40% (13.80%)
2020	19.31% (12.41%)	5.98% (7.10%)	19.20% (10.14%)	10.35% (7.90%)	25.61% (14.75%)	-11.98% (9.59%)

Table 3: Results (Annual Returns/Max Drawdown) for SSE Composite Index

	Random Forest		GBM		XgBoost	
Year	Whole	Hurst segment	Whole	Hurst segment	Whole	Hurst segment
2019	-10.43% (6.96%)	-6.00% (6.96%)	-10.36% (7.64%)	-6.49% (8.70%)	-8.84% (7.64%)	2.09% (4.91%)
2020	-3.88% (6.43%)	11.98% (10.54%)	9.54% (9.16%)	0.81% (12.76%)	-18.02% (6.67%)	-5.27% (8.49%)

Table 4: Results (Annual Returns/Max Drawdown) for Gold SPDR Shares

	Random Forest		GBM		XgBoost	
Year	Whole	Hurst segment	Whole	Hurst segment	Whole	Hurst segment
2019	3.42% (22.42%)	40.83% (21.32%)	-46.01% (48.36%)	-77.95% (39.83%)	-35.32% (48.25%)	-84.80% (55.32%)
2020	-55.10% (33.43%)	15.17% (14.70%)	-105.21% (116.44%)	-74.40% (37.29%)	-71.66% (36.34%)	-75.88% (29.73%)

Table 5: Results (Annual Returns/Max Drawdown) for Bitcoin-USD

The structure to build and select the models for all the asset classes is same. Above tables capture the annual return/maximum drawdown (on daily basis) for the trading scenario. Hurst Exponent based segmented approach modelling resulted in higher gain than no segmented approach in four out of six times for S&P 500 Index, and five out of six scenarios in Gold SPDR Shares. For Bitcoin, both approaches (Segmented model, and non-segmented model) had same win ratio of 3/3. For SSE Composite Index the performance was reversed i.e. Segmented model performed better than non-segmented model in only one scenario out of six. In most scenarios (for all assets), the segmented approach also helped reduce the maximum drawdown on daily basis compared to non-segmented approach. The codes for the project can be found at [24].

5. Discussions

Hurst Exponent based segmentation approach was more profitable than non-segmented approach for S&P 500, and Gold SPDR shares assets, and performed poorly for SSE Composite Index. For Bitcoin, both segmented and non-segmented approach performed equally, but Hurst Exponent based segmentation had a lower maximum drawdown. S&P 500, and Gold are both mature markets. SSE Composite is developing market, and Bitcoin in an emerging market. The models were built using data from 2011 to 2018 to train and validate the model, and score the selected model on 2019-2020 data. Markets change at dynamic pace, and increased money supply in 2020 by Fed, and COVID had a big impact on the markets. As a next step, a moving window model (model trained on recent data) could be explored to better gauge the effectiveness of the segmented approach. Also use of Deep Learning Models, and/or adding volatility to refine the trading strategy further could help improve the model performance.

Bibliography

[1] Hurst, H. (1951). Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers, 1, 519-543.
[2] Mandelbrot, B.B., & Van Ness, J. (1968). Fractional Brownian motions fractional noises and applications. SIAM Review, 10, 422-437. http://dx.doi.org/10.1137/1010093.
[3] Mandelbrot, B.B. (1982). The Fractal Geometry of Nature. New York: WH Freeman and Co.
[4] Peters, E. (1991). Chaos and Order in the Capital Markets: A new view of cycles prices and market volatility. John Wiley & Sons: New York.
[5] Lipka, J., & Los, C. (2002). Persistence characteristics of European stock indexes. Working Paper 2002 Kent State University Kent OH.
[6] Corazza, M., & Malliaris, A. G. (2002). Multifractality in Foreign Currency Markets. Multinational Finance Journal, 6, 387-401.
[7] Cajueiro, D.O., & Tabak, B. M. (2004). Ranking efficiency for emerging markets. Chaos Solitons & Fractals, 22, 349-352. http://dx.doi.org/10.1016/j.chaos.2004.02.005.
[8] Mitra, S. 2012. Is Hurst Exponent Value Useful in Forecasting Financial Time Series? Asian Social Science Vol. 8, No. 8.
[9] Katsev, S., and I. L'Heureux, Are Hurst exponents estimated from short or irregular time series meaningful?, Comput. Geosci., in press, 2003.
[10] Kirichenko, L.; Radivilova, T.; Deineko, Z. Comparative Analysis for Estimating of the Hurst Exponent for Stationary and Nonstationary Time Series. Inf. Technol. Knowl. 2011, 5, 371–388.
[11] Morales, R., T. Di Matteo, R. Gramatica, and T. Aste (2012). Dynamical hurst exponent as a tool to monitor unstable periods in financial time series. Physica A, accepted, Arxiv preprint arXiv:1109.0465.
[12] Pyo S, Lee J, Cha M, Jang H. Predictability of machine learning techniques to forecast the trends of market index prices: hypothesis testing for the Korean stock markets. PLoS One 2017 Nov;12(11):e0188107.
[13] Yakup Kara, Melek Acar Boyacioglu, and Ömer Kaan Baykan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. Expert systems with Applications, 38(5):5311–5319, 2011.
[14] Qiu M, Song Y. Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. PLOS ONE. 2016;11(5):1–11.
[15] Martinez, L., Guercio, M., Bariviera, A., and Terceno, A., 2016. The impact of the financial crisis on the long-range memory of European corporate bond and stock markets. Emprica, Journal of Applied Economics and Economic Policy, 1-19. http://dx.doi.org/10.1007/s10663-016-9340-8.
[16] A. Ammy-Driss, M. Garcin, Efficiency of the financial markets during the COVID-19 crisis: Time-varying parameters of fractional stable dynamics (2020) arXiv preprint arXiv:2007.10727.
[17] Bao, W., Yue, J. & Rao, Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one 12, e0180944 (2017).
[18] Song, Y., J.W. Lee, and J. Lee, A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction. Applied Intelligence, 2019. 49(3): p. 897-911.
[19] Dixon Matthew, Klabjan Diego, Bang Jin Hoon. Classification-based Financial Markets Prediction using Deep Neural Networks. 2016;2016 arXiv preprint arXiv:1603.08604.
[20] Islam, M.R.; Nguyen, N. Comparison of Financial Models for Stock Price Prediction. J. Risk Financial Manag. 2020, 13, 181.
[21] T. Chenoweth, Z. Obradovic, Embedding technical analysis into neural network based trading systems, Appl. Artif. Intell. 10 (1996) 523–541.
[22] Wang, Y. (2014) 'Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HSI', Int. J. Business Intelligence and Data Mining, Vol. 9, No. 2, pp.145–160.
[23] Qian, B., & Rasheed, K. (2004, November). Hurst exponent and financial market predictability. In IASTED conference on Financial Engineering and Applications (pp. 203-209).
[24] https://github.com/Sidhus234/WQU-Capstone-Project-2021

Disclaimer

Appendix

A. Resources:

Code repository: https://github.com/Sidhus234/Hurst-Exponent-Trading-Strategy

pandas_datareader.version = 0.8.1 https://pandas-datareader.readthedocs.io/en/latest/

pandas.version = 1.0.1 https://pandas.pydata.org/

numpy.version = 1.18.1 https://numpy.org/

matplotlib.version = 3.1.3 https://matplotlib.org/

yfinance.version = 0.1.54 https://pypi.org/project/yfinance/

scipy.version = 1.4.1 https://www.scipy.org/

hurst.version = 0.0.5 https://pypi.org/project/hurst/

sklearn.version = 0.22.1 https://scikit-learn.org/stable/

ta.version = 0.7.0 https://pypi.org/project/ta/

pickle: https://github.com/python/cpython/blob/3.9/Lib/pickle.py

os: https://docs.python.org/3/library/os.html