STOCK PREDICTIONS

Exxon Mobile Stocks Analysis

Abstract

This paper analyzes the predictive power of stock momentum paired with the weekly importation report of crude oil. A variety of Machine Learning models are used to identify patterns and form a predictive model. The end goal is to take the beginning of the day information and accurately predict the outcome for that day. The best algorithm to achieve this goal requires a reduction of features and an ensemble of several algorithms. It is concluded that it is possible to predict positive and negative trends by using momentum indicators coupled with crude oil logistical information, but it is not possible to know future profitability. (This is a study continuation.)

Bryce Perkins

brycej.perkins@gmail.com

The Dataset

The main dataset used for this study is a combination of datasets found using an online tool called Quandl. Quandl provides many different datasets relating to the financial industry. I was able to retrieve data for daily stock performances and weekly reports for crude oil importation. Premium datasets were also offered, but not explored. Exxon Mobile was singled out as the company of study.

This information is regularly viewed by investors, competition, and news broadcasting companies. It is used to evaluate company worth and can be used to predict economic trends. Many academic support the efficient-market theory that states that stock prices include that all past and present information as well as some degree of future information about the underlying firm (Tonkinson, Hale, Perkins, & Gardner, 2017). On the other side of this argument, it has been revealed that there are many inconsistencies that present incorrect prices. Investors that can identify these inconsistencies can exploit them for profit.

The Problem

The main goal in this study is to continue the efforts in identifying mispricing related to momentum, and, the furthering exploration, logistical information relating to the company's sector.

The core function of this model is to be able to predict the future performance of a stock based on any current information available and past performance. Through supervised training, it will classify an expected increase or decrease on day to day instances.

This study is a continuation of a recent study, Momentum Stock Predictor (Tonkinson, Hale, Perkins, & Gardner, 2017). In this previous study, it was found that using full historical stock information generally returned poor results, but isolated ranges increased performance. This may be due to industry changes, public opinion, etc. Trends within a market or company may last only certain periods of time. This can skew results for current trends, which is what an investor is typically focused on. If one of these trends is identified, it may prove to be more beneficial to tailor a model to only use related data.

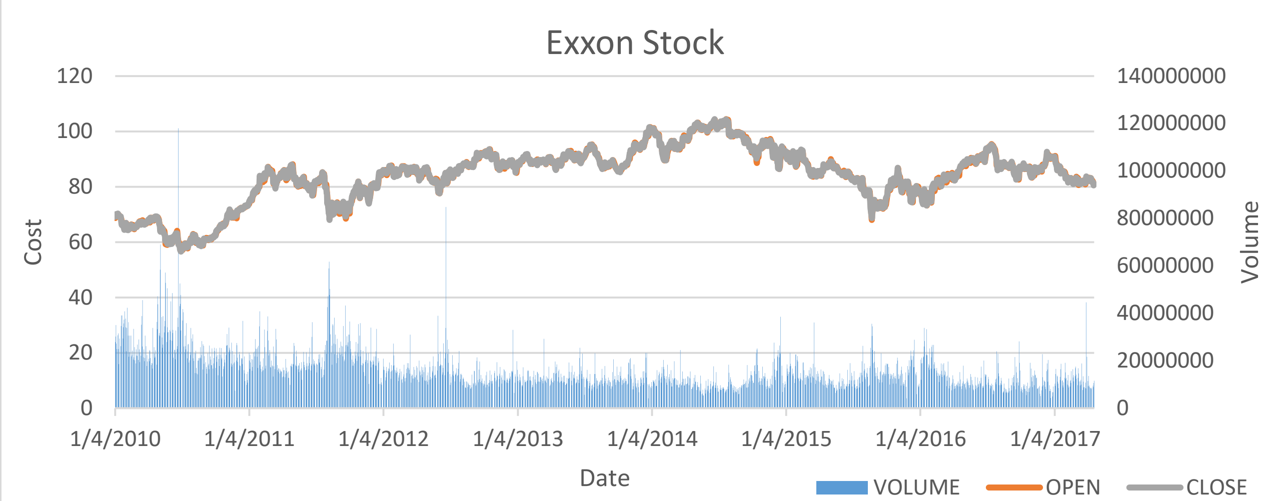
Results from this previous study showed that momentum can increase the predictability of stock, but not enough to provide a comfortable base for investing. The study focused on large tech companies and may only be relevant to such.

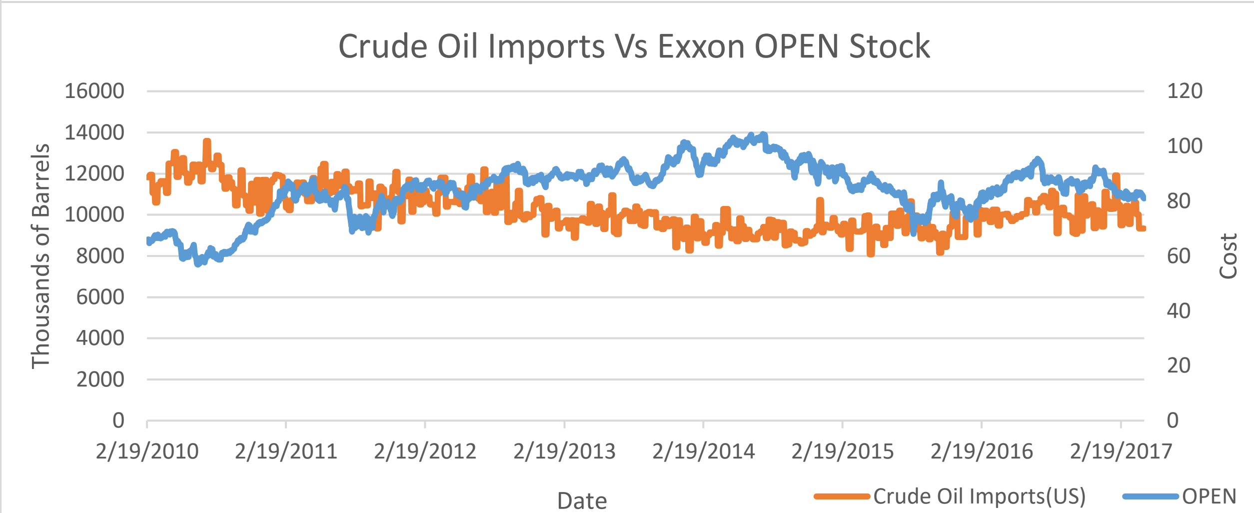
This continued study will branch out with the hypothesis that momentum can be used as a good indicator for stock trends and incorporate additional sets of data. One problem that we encountered working with tech companies was they had fewer publicly accessible datasets in relation to their sector. I chose to explore the sector dealing with oil, Exxon Mobile in particular, as information related to the movement, storage, and processing of products was much more accessible.

Exploration

The initial combined dataset held the fields: Date, Open, High, Low, Close Volume, Ex-Dividend, Split Ratio, Adj. Open, Adj. High, Adj. Low, Adj. Close, Adj. Volume, Weekly U.S. Imports of Crude Oil and Petroleum Products, and over 20 other subcategories, parsing out the import total.

In examining the data it is noticeable that volume spikes are associated with relatively dramatic changes in stock price, but the spikes are not indicative of a positive or negative change. Much to the same effect, crude oil spikes, though much smaller, seem to denote a change in stock price. Typically, crude oil upward spikes appear to be correlated with decreases in stock price.





Since historical stock information is on a business day to day status and the Crude Oil imports is a weekly report, each instance was set to refer back to the nearest importation report. Also, the stock information must refer to the previous day's volume and the Close attribute was removed as the model will not have that information when predicting the outcome of the day's performance.

Just as we found in our previous study, using more recent subsets of historical data proved to be more beneficial when trying to extract a predictive model. I restricted the historical data to look back only six years instead of the full stock history. The final instance count came out to be 1805 with a 52.355% positive day-to-day trend.

Finalized Features:

1. Date – Day of the trade.
2. Open – Opening price of the stock.
3. Momentum5 – The average difference over the last 5 days.
4. Momentum10 – The average difference over the last 10 days.
5. Momentum30 – The average difference over the last 30 days.
6. USCrudeOilImports – The amount (by thousands) of barrels imported into the US for the previous week.
7. Volume – The number of stock that were traded.

Technical Description

Much of the approach in this study is based on experience from the last study. A logistic model proved to be the most accurate model across all tests, but when applied to a different sector, dealing with oil, did not maintain the same performance. Alternatives were explored, both by themselves as well as in ensembles.

Many different algorithms were used in this study in order to tailor an ensemble to best fit the data. The three main accurate algorithms explored were OneR, RandomForest, and LinearNNSearch.

OneR is a simple classifier that focuses on creating one rule for each attribute. It then calculates the total error for the rules of each attribute. It finally chose the attribute with the least amount of error to be the classifier.

RandomForest implements many random trees and lets them vote on the classification. Each random tree is built off a subset of the training set. Using this data, branches are made by selecting random features and comparing likelihoods. The best split is made and the tree is grown to the largest extent. Restrictions can be made on how big a tree can grow as well as how many instances are required in a branch.

IBK (Nearest Neighbor) is an instance based learner that classifies by comparing similar instances or neighbors. The nearest neighbor, or similar instances, vote on the classification of the tested instance.

Overfitting was not entirely relevant until I tested tree-like algorithms and noticed that an unrestricted algorithm could create over a thousand branches to describe a dataset of only 1805 instances. This meant that many branches were created to accommodate only one instance of data, a clear sign of over-fitting. To overcome this issue I was able to restrict the amount of branches created and require that a branch represent more than one instance of data. This decreased overall accuracy, but led to the awareness that I need to manage overfitting the data.

The partitioning of data was split a couple different ways in an effort to increase accuracy as 1805 instances is a low amount of data. Initial trainings were performed using cross validation and having a separate file for testing instance. This achieved the highest accuracy. Another method I thought would perform well was to partition the data 90:10 while maintaining the order of the data by date. I thought this might allow for a trend to be learned and predicted, but, while it was better than the base trend, it did not achieve an acceptable accuracy. All test sets were subsets of the trading history referring back 15 to 30 days from the day of the test. This was done to establish investment scenarios and test the abilities of the model.

Analysis

The best results found were through an ensemble of algorithms voting for the prediction. The algorithms included were: Random Forest, IBK, and OneR. The overall accuracy was 58.4488%, with a RMSE of .5225.

The problem/challenge is far from being solved as new technologies, social movements, crowd movements, natural occurrences…, all add substantial amounts of noise. While some of these events are predictable, most happen almost randomly and the effects are close to instantaneous and volatile.

While the predictive capabilities of this final model are better than the 52% positive trend that Exxon Mobile’s stock has demonstrated, this tool should be used cautiously and in ensembles with attentive knowledge of the stock market and awareness of the events currently surrounding the targeted stock. Because of the narrowness of the stock selection it is not recommended to use this model (I would not trust this with actual money. For the majority of the test investment scenarios I lost money). Also, the fact that the types of algorithms that were used to create final models were very different between the previous study and this one suggests that tailoring to each company/sector may be required.

Future testing will attempt to isolate some of the seemingly unpredictable variables and incorporate them into the existing model. Public opinion of a company or sector can greatly influence the trading of a company's stock, as seen in the recent event involving a man being lawfully removed from plane. Continued exploration of the Quandl datasets has shown there to be additional points of interest within the observations of oil flow and storage. The JODI dataset makes monthly measurements of oil storage level in various regions of the world. This could be relevant to the trading of oil related stocks as a rise in storage could suggest a stagnation in the market. Lastly, I would like to incorporate the prediction of the change in amount of the targeted stock. This would allow investors to determine if the cost of transaction would be covered by the difference if a trade was made.

References

Tonkinson, C., Hale, K., Perkins, B., & Gardner, B. (2017). *Momentum Stock Picker.* Provo: BYU.