Stock Market Prediction

Mark Dunne

March 17, 2015

Abstract

In this report we analyse existing, and explore new methods of stock market prediction. We take three different approaches at the problem; Fundamental analysis, Technical Analysis, and the application of Machine Learning. We find evidence in support of the weak form of the Efficient Market hypothesis, that the historic price does not contain useful information but out of sample data may be predictive. We show that Fundamental Analysis and Machine Learning could be used to guide an investors decisions. We demonstrate a common flaw in the methodology of Technical Analysis practitioners and show that it produces limited useful information. Based on our findings, an algorithmic trading algorithm is developed and entered into the Quantopian trading competition.

Contents

Chapter 1

Introduction

My project was dealing with stock market prediction.

This project was chosen because of the purity of the data analytics challenge it poses.

about the stock market

how many people do this all the time

1.1   Project Focus

The project and report is primarily a data analytics project and so will focus on the accuracy of the prediction of the stock market, rather than trying to profit from it. Although later when we will build on top of predictive models to attempt to profit from them, this is not the focus of the project.

Chapter 2

Data

For this project, we chose the Dow Jones and its components as a representitive bundle of stocks. The dow jones is a large index traded on the New York stock exchange. It is a prices-weighted index over 30 component companies [todo show calculation]. All companies in the index are large publically traded companies, leaders in each of their own sectors. The index covers many different sectors featuring companies such as Microsoft, Visa, Boeing, and Walt Disney.

We wanted to use a set of companies already picked by someone else so that we don't open ourselves to methodology errors / fishing expeditions to find a set of companies that our algorithms do happen to work for.

The dow jones was chosen because it is well known, and has a relatively small number of compontents when compared to indices such as the S&P 500 which has over 500 components at the time of writing.

This small but representitiive set allowed for a managable dataset given limited resources. Although there were only 30 companies, there was no lack of data to study. To test many of the hypothesis laid out in this report, we were able to extract many datasets, the smallest containing [todo find smallest num examples] examples.

2.1   Data Sources

Used quandl

Chapter 3

Considerations in approaching the problem

Throughout the project, there are a couple of things that should be kept in mind. All three of these ideas, in their own way, emplore us to keep an open mind in that we might not actually find a profitable way to predict market movements.

3.1 Random Walk Hypothesis

The random walk hypothesis sets out the bleakest view of the predictibility of the stock market. The hypothesis says that the market price of a stock is essentially random. The hypothesis implies that any attempt to predict the stock market will inevitablyy fail.

The term was popularized by [todo find sources]. He compared the stock market to a coin flip and asked students to try and predict the movements of the coin flip. some believed they could.

3.1.1 Qualitative Visual Similarity to Random pattern

The stock market certainly looks random to the eye of any human observer. To demonstrate this, we generated a random process with similar visual characteristics to the dow jones index

[todo outline generations functions]

[todo plot graphs]

3.1.2 Quantitative Difference to Random pattern

The above section gives us pause when attempting to predict the stock market. If the market is random, then there is nothing to predict. This would leave

nothing to do for the rest of the project. However, it can be demonstrated that stock data is indeed fundamentally different to random data.

[todo plot random difference]

On every day from the year 2000 to 2014, we sysimulated an investment on the dow jones index. We then counted the number of days it took for the investment to gain or lose 5% of its original value. When it lost 5% of its value, it was put into the red set, when it gained 5% of its original value, it was put into the green set. The graph shows 2 overlaid histograms detailing how long it took for an investment to lose of gain 5%.

What this graph shows is that the market generally creeps upwards but is prone to sudden drops downwards.

This demonstrates that the stock market is fundamentally different to random data. This gives us hope for the remainder of the project. If the market price is not random, then it might be worth investigating and trying to predict.

3.2   Efficient market hypothesis

Broadly, the efficient market says that the market is very efficient at correctly pricing products.

It comes in three flavors

Weak-form Efficient Market Hypothesis   The weak form of the hypothesis says that no one can profit from the stock market by looking at trends and patterns within the price of a product itself. It is important to note that this does not rule out profiting from predictions of the price of a product based on data external to the price. We will see examples of prediction based on both in sample and out of sample data, and provide evidence in support of the weak form.

Semi-Strong Efficient Market Hypothesis   The Semi strong form rules out all methods of prediction, except for insider trading. This means that if we are only to use public domain information in our prediction attempt, the Semi-Strong form says that we will be unsuccessful. Later in the project, we will provide results are seem to be inline with this hypothesis but not as good as with the weak form.

Strong form Efficient Market Hypothesis   The strong form says that no one can profit from predicting the market, not even insider traders.

Clearly, if we are to predict the stock market using only public information, we must hope that at most the weak form of the efficient market hypothesis is true so that at least then we can use external data to predict the price of a product.

3.3   Self Defeating Success

Finally there is the idea of a successful predictive model ultimately leading to its own dimize.

The idea here is that if there is a simple predictive model that anyone could find and apply for themselves, then over time all of the advantage will be traded and erroded away.

This is the same reason for the lack of academic papers on the topic of successfully predicting the market. If a successful model was made widely known, then it wouldn't take long until everyone tries to use it and the pattern ceases to exist.

3.4   Conclusions

The three preceding ideas ask us to keep an open mind on stock market prediction. It is possible that we will not be able to do it profitably.

Chapter 4

Attacking the problem - Fundamental Analysis

The first approach we take at solving the problem of market prediction is to use Fundamental Analysis. This approach tries to find the true value of a company, and thus determine how much one share of that company should really be worth. The assumption then is that given enough time, the market will generally agree with your prediction and move to correct its error. If you determine the market has undervalued a company, then the market price should rise to correct this inefficiency, and conversely fall to correct the price of an overvalued company.

It should be noted that Fundamental Analysis is compatible with the weak form of the efficient market hypothesis. As explained earlier, the weak form does not rule out prediction from data sources external to the price, which is what we will use to determine our fair market price.

Fundamental Analysis attempts to predict long term price movements, i.e on a year to year basis. [todo expand]

4.1 Fundamental Analysis limitations

There is an obvious pattern with fundamental Analysis. We are trying to find the quantify the true value of a company when almost every company has in some way or another some purely qualitative value

Fundamental Analysis methods does not attempt to capture, and so it difficult to build a software solution to do so. This leaves a large gap in knowledge an algorithm could learn about a company. How should it quantify the value of a brand, the size of its customer base, or a competitive advantage?

These are three examples of some of the many things that a human investor might take into account when deciding who to invest in, but they are untouchable within the scope of this project.

Instead, we are limited to purely quantitative company metrics. We will look at two of the most common metrics, Price to Earnings ratio and Price to

Book ratio.

4.2   Price to Earnings ratio

The first metric for the value of a company that we will look at is the Price to Earnings ratio. The price to earnings ratio is calculated as

[todo show pe ratio formula]

Informally, what this calculates is the price an investor is willing to pay for every \$1 of company earnings. If this ratio is high, it might translate to high investor confidence. If investor confidence is high, that might mean they expect high returns in the following year. There should then be a relationship between high P/E ratio and high returns in the following year.

To investigate this relationship, we plotted the P/E ratio for of 450 companies on the 31st of December against the change in stock price for the following year. We gathered these data points from the year 2000 to 2014. Below is a graph of this relationship.

[todo pe graph]

The best fit line was calculated using the standard Least Squares method. If the P/E ratio was indeed predictive, we might have expected a steeper slope in the best fit line, but we can see that there is a very weak correlation at best. It should also be noted that the more we remove outliers, the lower the slope becomes. This indicates that the the line is probably being pulled up by outliers rather than an actual correlation in the data.

We can also see that the vast majority of the data points are concentrated in one area, without any obvious pattern. It is clear that this data should not be used for prediction.

4.3   Price to Book ratio

The second metric for the value of a company that we will look at is the Price to Book ratio. The price to Book ratio is calculated as

[todo show pb ratio formula]

Informally, what this calculates is the ratio between the value of a company according to the market and the value of the company on paper. If the ratio is high, this might be a signal that the market has overvalued a company and the price may fall over time. Conversely if the ratio is low, that may signal that the market has undervalued the company and the price may rise over time. There should then be a relationship between high P/B ratio and low returns in the following year.

To investigate this relationship, we plotted the P/E ratio for of 350 companies on the 31st of December against the change in stock price for the following year. We gathered these data points from the year 2000 to 2014. Below is a graph of this relationship.

[todo show pb graph]

The best fit line was calculated using the standard Least Squares method. Although slope of the best fit line is greater than that of the P/E ratio, this is the oposite of what we might have predicted. The data suggests that a high P/B ratio is mildly [todo better word] predictive of a high growth in the stock price. Although this may be counter to what literature might suggest, it the data itself may not be an error. A high P/B ratio could be a better signal of investor confidence than what P/E ratio was and so we can carry over the logic from P/E ratio and apply it to the P/B ratio.

However, as previously mentioned this finding runs directly apposed to the literature on the subject [todo backup]. We must therefore be somewhat sceptical of using this feature for prediction.

4.4   Conclustion

We evaluated two Fundamental Analysis metrics for their predictive value, and found only tenuous relationships.

These predictions are also very long term, looking one year into the future. Predictions on this time scale were not the focus of the project, instead we wanted to focus on predicting daily trends in the market.

It is because of these issues that we moved on from Fundamental Analysis.

Chapter 5

Technical Analysis