Stock Market Prediction

Mark Dunne

January 18, 2015

Contents

Abstract

1 Introduction

1.1 Introduction

1.2 Project motivation

1.3 Importance and prevalence of the problem

Explain size of the financial industry affected by the problem and why having better tools to predict the stock market may be important to the average person

1.4 Project goals

1.5 Project outline

How the project will meet those goals

2 Background

2.1 The Stock Market

2.1.1 What is the stock market

2.1.2 How the stock market works

2.1.3 Components of the stock market

2.1.4 Terminology

2.1.5 The Efficient Market Hypothesis

Throughout this area of investigation, there is a very large elephant in the room, the Efficient Market Hypothesis (EMH). The hypothesis talks specif-

ically about an agents ability to profit from make inefficiencies, i.e when stocks and shares are mispriced by the market. Its strongest proponents would claim that the very title of this report, predicting stock market, is all but impossible. The EMH comes in three main forms.

- The weak form of the efficient market hypothesis claims that prices fully reflect the information implicit in the sequence of past prices.
- The semi-strong form of the hypothesis asserts that prices reflect all relevant information that is publicly available
- the strong form of market efficiency asserts information that is known to any participant is reflected in market prices.

[2]

Informally, the weak form implies that you cannot profit using strategies shaped on historic data, the semi-strong form implies that there is profit only to be made from insider trading, and the strong form says that even this is futile. Clearly if this project is to be of any success, we must hope that the hypothesis is wrong and does allow for a sufficiently intelligent agent to profit.

Luckily, many researchers do indeed question the validity of the hypothesis. There is evidence that the stock market does not always follow EMH. [1, (Basu 1983)] showed that certain analysis could yeild information useful in future market forecasts.

2.2 Analysis of the problem

2.2.1 Explanation of the difficulty of the problem

2.2.2 Separation of profitability and accuracy

2.2.3 Temporal reach of prediction

2.2.4 Formal definition of the problem

2.3 Review of existing work

3 Methodology and Data

3.1 Tools Used

3.1.1 Python, Numpy, Pandas

3.1.2 Quantopian/Zipline and Pyalgotrade

3.1.3 Statsmodels

3.2 Data Used

3.2.1 Data sources

3.2.2 Format of the data

3.2.3 Adjusted prices

3.3 Simulation of strategies

Similarity to real life

3.4 Defining a successful model

Statistical significance of a model

4 Attacking the problem - Fundamental Analysis

We begin by approaching the problem using Fundamental Analysis. Fundamental Analysis of stocks and shares is the earliest and simpliest form of prediction.

4.1 PE Ratio

5 Attacking the problem - Technical Analysis

5.1 Hobbyist Approaches

5.2 Review of Metrics

5.3 OLMAR algorithm

5.4 StatsModels

6 Attacking the problem - Machine Learning

6.1 KNN on metrics

[3]

References

[1] Sanjoy Basu. The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence. *Journal of financial economics*, 12(1):129–156, 1983.

[2] Elroy Dimson and Massoud Mussavian. A brief history of market efficiency. *European financial management*, 4(1):91–103, 1998.

[3] pybrain.org. Classification with feed-forward neural networks, January 2015.