A Short-Term Algorithmic Trading Model for the US Stock Market

By Emmad Arif

Advisor: Professor James Glenn

Introduction

In this project, I plan to develop and present a comprehensive model to trade certain stocks in the US equities market.

Motivation

Most strategies used by beginner traders today are based purely on fundamental analysis i.e. traders evaluate a security by examining the company’s balance sheet health, debt, recent earnings etc. and decide whether to open/close positions accordingly. Such strategies have proven to be profitable in the long run, however they break down if employed on a shorter time horizon, due to market fluctuations not captured by fundamental analysis.

Methodology

I propose to develop a comprehensive model which is threefold in its approach and incorporates a wider variety of data, in an attempt to capture a fuller snapshot of the market.

My primary data sources will be historical market data (price and volume). Since our goal is to present a strategy which is successful in the short run, it makes sense to evaluate recent price action in light of historical trends and use it predict future behavior. I plan to approach this as a supervised learning problem, where the inputs will be metrics derived from the price and volume data. These metrics may include moving averages, price momentum etc.

While historic pricing information is certainly an important factor in predicting future price movements, at the end of the day, liquidity is a variable which can only be predicted by analyzing how confident traders feel about the market situation. For instance, if most traders believe the price for a certain stock will rise soon, we can reasonably expect a significant rise in demand for the stock. This is a self-fulfilling prophecy, since a higher demand leads to higher prices.

Moreover, public remarks made by certain influential personalities have a marked impact on stock prices. As an example, Elon Musk’s tweet on August 7, 2018 regarding potentially “taking Tesla private at $420” led to a 14% rise in Tesla stock price over a few hours. Similarly, President Trump’s tweets on a potential trade deal with China led to a noticeable impact on major indices such as the Dow Jones Industrial Average.
My **secondary data sources** incorporate this “human element” of trading into the model. I plan to analyze recent tweets, Facebook posts, and online news articles related to the company, and conduct sentiment analysis on this data. There are a number of text analysis packages available, including *textblob*, which I will use in conjunction with media websites’ APIs to collect and analyze information. The information drawn from this analysis will enrich my model, since it will help me emulate human reaction to price movements more accurately.

Lastly, I will analyze some **tertiary data sources**, which will present valuable information about the state of the economy at large, as well as the relative performance of a certain sector in the stock market. Examples of such data sources include the US GDP growth rate, trade deficits, oil prices etc.

Eventually, I will incorporate various elements from each category into the model, depending on their success rate in back tests. My final report will present the results of the analysis conducted, as well as a description of the final model(s), along with metrics measuring its success rate in different conditions.

Moreover, I will also provide an application (and corresponding documentation) which will allow users to tweak the variables in my model, run it in real-time on any stock, as well as automatically back test it. In doing so, I hope to make the model accessible to non-technical persons, so that they may employ the strategies presented confidently, based on back test results.

**Stages of Development**

1. Collection of Historic Pricing and Volume Data
2. Develop a Multivariate Model based on variables derived from:
   a. Primary Sources: Recent Pricing and Volume Data (EMA, MACD etc.)
   b. Secondary Sources: Public Sentiment Surrounding the Company (sentiment analysis)
   c. Tertiary Sources: State of the Economy, Political Stability metrics etc.
3. Conduct Back Testing and Calculate Success Rate
4. Refine Model to Maximize Success Rate

**Checkpoints**

1. 4 February – 17 February (2 weeks) [Data Collection & Research]:
   - Collect and clean historical market data.
   - Research techniques to analyze time-series data.
2. 18 February – 17 March (4 weeks) [Primary Sources]:
   - Conduct time series analysis on market data.
Develop preliminary forecast model.

3. 18 March – 14 April (4 weeks) [Secondary and Tertiary Sources]:
   Conduct sentiment analysis on data from news/social media and link it to the model.
   Add tertiary inputs to model.

4. 15 April – 28 April (2 weeks) [Finalize Deliverables]:
   Write up final report.
   Finalize application enabling user to apply model to any stock dynamically.

**Deliverables**

1. Source Code:
   a. R/Python scripts written to conduct statistical analysis and create final model.
   b. An application (and documentation) which allows the user to:
      i. Run final model on any stock in real-time and get a buy/hold/sell signal,
      ii. Back test final model on any stock for a desired time-frame and determine success rate.

2. Final Report:
   A report summarizing the data and methodology used to conduct the analysis, a description of significant variables and results of the model created.