Abstract

By studying the fluctuations of stocks we can try and answer the following question: Is there a pattern in the stock price trends in December and January that one could use to decide whether to buy/sell/keep a stock? Specifically, the analyzing of trends of the adjusted closed prices across December and January by using the MSE (minimum mean square error) to compare the trend line of each stock price for several large companies.

Stock Price Trend Analysis with MSE

Claire Gong

December 2020

Contents

1 Introduction

In general, predicting the prices of stocks is not reliable. However, there may be some trends among the stock fluctuations. Specifically, at the end of the year, people may want to change their investment strategies, which may result in the selling of stocks in December/January. I have created this project in hopes of validating the hypothesis with the data of stocks from several companies across several years.

2 Formulation

2.1 MSE solution for linear equation

The stock prices will be modeled by observations across time that are established through numerous points of dates and adjusted closed prices (N):

$$(t_1, \hat{y}_1)(t_2, \hat{y}_2)\dots(t_N, \hat{y}_N)$$

where $\hat{y}_t$ is the observed adjusted closed price at time t ($1 \leq t \leq N$).

To determine if the stock has increased or decreased, we want to create a line of best fit (linear equation) of all of the points using the equation MSE (minimum mean square error).

$$y_t = m \cdot t + b \tag{1}$$

We need to find m and b such that the MSE equation is minimized.

$$\text{MSE} = \min_{m,b} \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2$$

In order to solve for m and b, we need to derive the MSE equation versus both m and b.

2.1.1 Derivative for m

$$\frac{\partial \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{\partial m} = \frac{1}{N} \sum_{t=1}^{N} 2(y_t - \hat{y}_t) \frac{dy_t}{dm} = \frac{1}{N} \sum_{t=1}^{N} 2(y_t - \hat{y}_t) \times t \tag{2}$$

2.1.2 Derivative for b

$$\frac{\partial \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{\partial b} = \frac{1}{N} \sum_{t=1}^{N} 2(y_t - \hat{y}_t) \frac{dy_t}{db} = \frac{1}{N} \sum_{t=1}^{N} 2(y_t - \hat{y}_t) \tag{3}$$

2.1.3 Solving for m and b

Setting EQ-2 and EQ-3 to 0. We first need to rearrange m and isolate b in their respective equations in order to do substitution. Here is m rearranged:

$$m \sum_{t=1}^{N} t^2 - \sum_{t=1}^{N} \hat{y}_t t + b \sum_{t=1}^{N} t = 0 \tag{4}$$

Here is b isolated:

$$b = -\frac{m}{N} \sum_{t=1}^{N} t + \frac{1}{N} \sum_{t=1}^{N} \hat{y}_t \tag{5}$$

Substituting b for m (see appendix):

$$m\left[\sum_{t=1}^{N} t^2 - \frac{1}{N}\left(\sum_{t=1}^{N} t\right)^2\right] = \sum_{t=1}^{N} \hat{y}_t t - \frac{1}{N}\sum_{t=1}^{N} \hat{y}_t \sum_{t=1}^{N} t \tag{6}$$

Final equation of m isolated:

$$m = \frac{\sum_{t=1}^{N} \hat{y}_t t - \frac{1}{N}\sum_{t=1}^{N} \hat{y}_t \sum_{t=1}^{N} t}{\sum_{t=1}^{N} t^2 - \frac{1}{N}\left(\sum_{t=1}^{N} t\right)^2} \tag{7}$$

Final equation of b isolated. Substituting m EQ-7 into b EQ-8:

$$b = -\frac{\sum_{t=1}^{N} t\left(\sum_{t=1}^{N} \hat{y}_t t - \frac{1}{N}\sum_{t=1}^{N} \hat{y}_t \sum_{t=1}^{N} t\right)}{N\sum_{t=1}^{N} t^2 - \left(\sum_{t=1}^{N} t\right)^2} + \frac{1}{N}\sum_{t=1}^{N} \hat{y}_t \tag{8}$$

2.2 Normalization of MSE solution

In order to focus on the relative change between the stocks between the given time period, we need to normalize the change of the stock prices. This cancels out the absolute difference across the stocks as well. While directly comparing the slope across different companies, one may run into intensive cognitive load. To combat this, normalization is required. Let P_B be the price of the beginning time and P_E be the price of the end time.

$$\Delta P = P_E - P_B \tag{9}$$

In order to normalize ΔP, we need to divide it by P_B.

$$\%\Delta P = \frac{P_E - P_B}{P_B} = \frac{P_E}{P_B} - 1 \tag{10}$$

Referring to EQ-1, we can substitute in EQ-1 into EQ-10. Note that P_B is equal to b (the y-intercept), P_E is equal to y, and N is equal to t.

$$\%\Delta P = \frac{P_E}{P_B} - 1 = \left[\frac{y}{b} - 1\right] = \frac{mt}{b} = \frac{mN}{b} \tag{11}$$

3 Python program

3.1 Python code outline

The python code is made to calculate:

1. For one stock:
   - DEC stock gain

- JAN stock gain
- the difference

2. Do (1) and average for last 5 years (function call)

3. Do (2) and average for 10 stocks

Please visit my GitHub for the python implementation.

3.2 Python code packages

Python packages used to help:

pandas Used to analyze the data of adjusted closed prices of the stocks from several companies.

sys Used to read and write files.

yfiannce Historical market data from Yahoo! finance.

matplotlib For plotting graphs

4 Data

The adjusted closed prices of public traded companies are downloaded from the python package, yfinance, which uses historical market data from Yahoo! finance.

5 Experimental results

The output of the Python code includes scatter plots depicting the normalized slope values determined by MSE from December to January of the given year. Let the $\%\Delta P(\cdot)$ be the normalized slope of month, where $\cdot$ is either December or January. In each specific quadrant of the table below, is the condition and hypothetical action.

$\%\ \Delta P(Dec)\downarrow$, $\%\Delta P(Jan)\uparrow$ buy	$\%\ \Delta P(Dec)\uparrow$, $\%\Delta P(Jan)\uparrow$ no change
$\%\ \Delta P(Dec)\downarrow$,$\%\ \Delta P(Jan)\downarrow$ sell	$\%\ \Delta P(Dec)\uparrow$, $\%\Delta P(Jan)\downarrow$ sell

Table 1: What to do with the December and January values

For example, see Fig-3 for Microsoft's stock performance of December and January, the points are mainly in the first and fourth quadrant. This means that the user should hypothetically either choose to buy/sell/keep (no change) or sell their stocks, depending on the user's risk tolerance.

5.1 Plot of two consecutive months

Here, we have two graphs showing the actual stock price data and the line of best fit, which shows the trend of the stock (slope, y-intercept) found from EQ-7 and EQ-8 across each month.

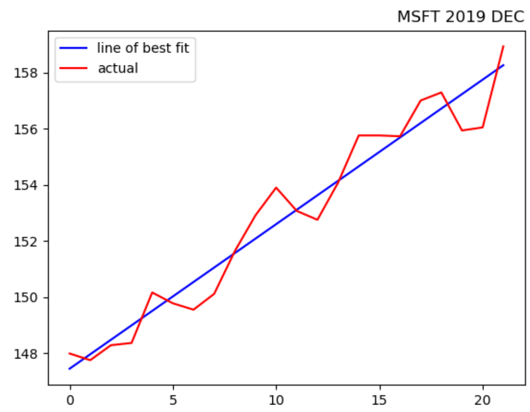

Figure 1: Microsoft December 2019 stock

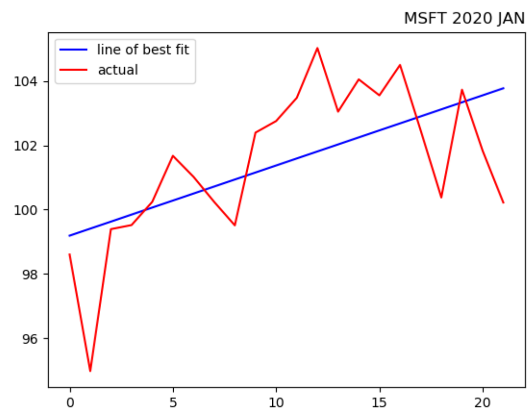

Figure 2: Microsoft January 2020 stock

5.2 Scatter plot for MSFT

The slope of best fits estsablished from EQ-7 (shown individually in Fig-1 and Fig-2) across 10 years are plotted in the graph below.

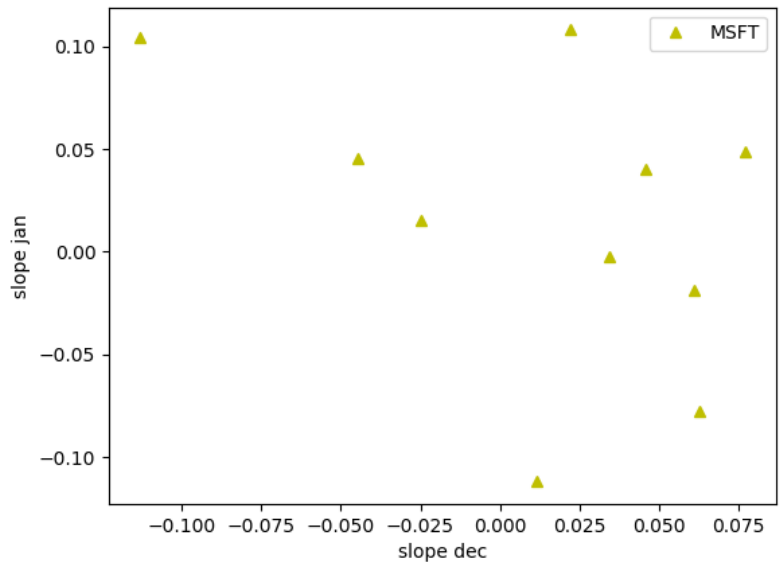

Figure 3: Microsoft scatter plot 2010-2020

5.3 Other scatter plots

I have also created scatter plots for several other companies including APPL, IBM, WLM, GOOGL, and AMZN. Please see references for Fig-4 through Fig-8.

6 Conclusion

This study's main question was: Is there a pattern in the stock price trends in December and January that one could use to decide whether to buy/sell/keep a stock? This was determined by looking at the slopes for each companies' stock prices during December and January with the minimum mean square error equation. Sadly, we did not observe any distinct patterns in the scatter plot data that one can clearly rely on. However, we did notice that most of the time, if the December plot is negative, most likely the January plot will be positive, meaning that one may choose to invest in December to receive some gains.

7 Appendix

An extra step leading to EQ-6

$$m\sum_{t=1}^{N}t^2-\sum_{t=1}^{N}\hat{y}_t t+\left(-\frac{m}{N}\sum_{t=1}^{N}t+\frac{1}{N}\sum_{t=1}^{N}\hat{y}_t\right)\sum_{t=1}^{N}t=0$$

References

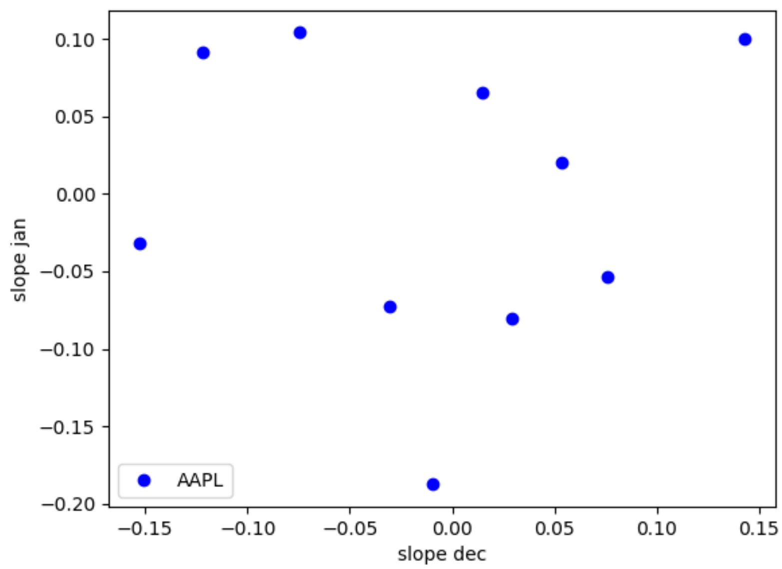

Figure 4: Apple scatter plot 2010-2020

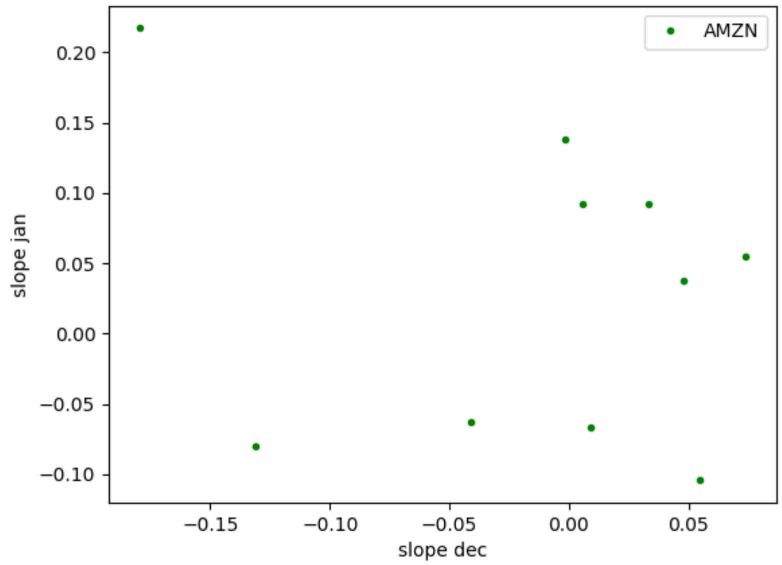

Figure 5: Amazon scatter plot 2010-2020

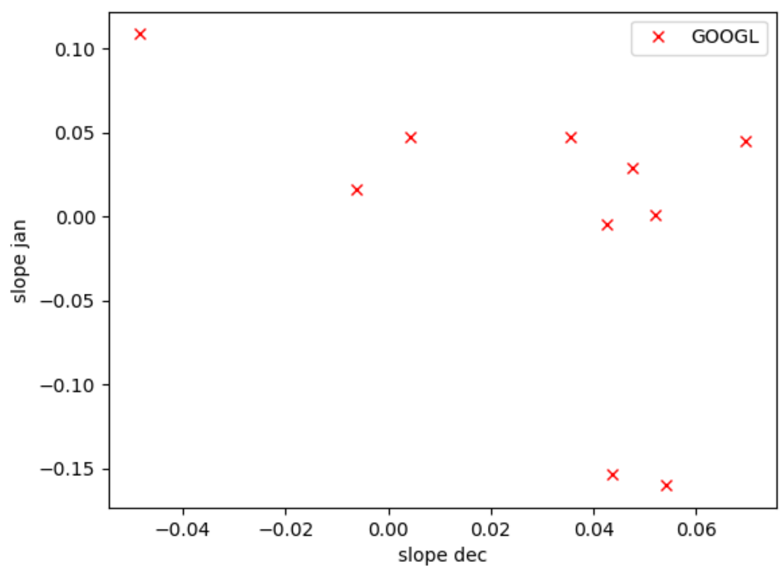

Figure 6: Google scatter plot 2010-2020

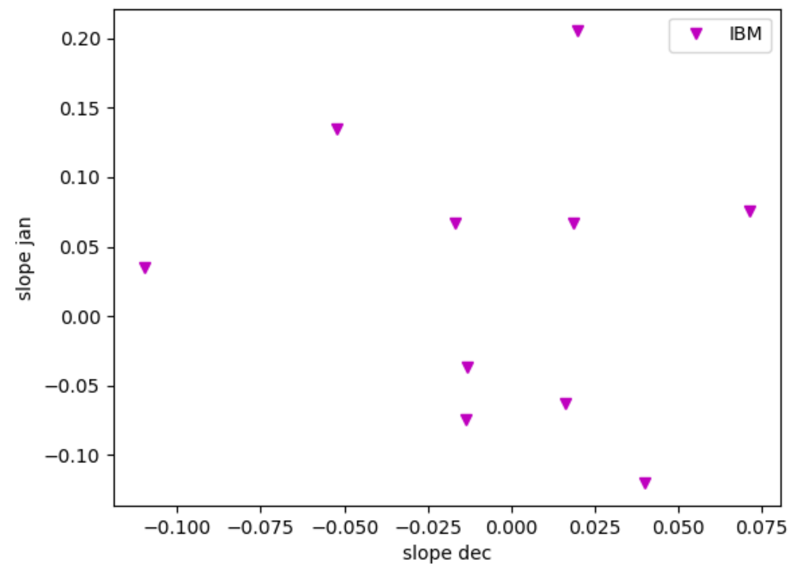

Figure 7: IBM scatter plot 2010-2020

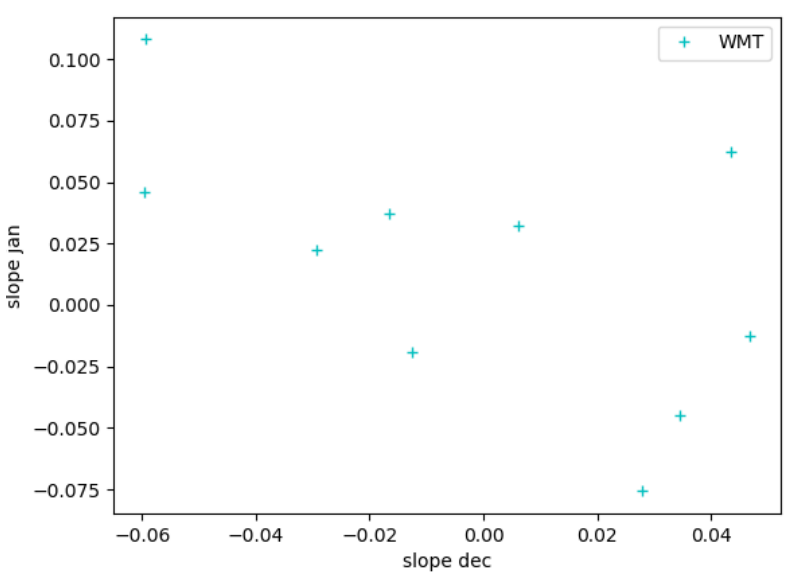

Figure 8: Walmart scatter plot 2010-2020