CS 131 MP2: Analysis of stock market prices as a function of previous price, volume, and momentum

Jonathan Cancio
Oliver Atienza
Justin Balderas
Reyster Fresco
Yna Ojeda
Ethan Tan
Marc Teves
Troy Valdez

May 25, 2018

1 Abstract

Stock prices seem to be unpredictable. Predicting the future price of shares in a particular company will allow stock traders to make more profitable trading strategies. The group tried to predict the closing price of seven different companies on 24 May 2018. The group performed a linear regression on their closing price, trading volume, and closing price momentum over varying timespans. The prediction error was found to be (to be continued...)

2 Introduction

Stock prices are unpredictable because there are simply too many factors that can influence the price of shares. While it would not be impossible to gather all the possible factors that could affect the price of a specific stock and make a mathematical equation describing how each factors into predicting the future price of a stock, it is likely that there are too many factors for any human built system to track.

The most obvious way of making money in the stock market via trading stocks is *arbitrage*, more commonly known as "buy low, sell high". Less known to the layman is the more effective technique called *momentum investing*, summarized as "buy high, sell higher".

Momentum investing became a popular strategy after investors and economists noticed that if a stock has been performing well over a time period of three to twelve months, then it will continue to do well. That is, if the price of a stock has had high returns for a long time, then it will continue to have high returns. On the other hand, if a stock has had poor returns for a long time, then it will continue to have low returns.

This tendency to stay increasing or decreasing is known as *momentum*, and can be quantified as the derivative of stock price at a specific point in time.

$$m_t = \frac{\mathrm{d}P(t)}{\mathrm{d}t}$$

where $P(t)$ is the price at time t.

If momentum investing is a good strategy, then there is some function

$$P(t, m)$$

that reasonably predicts $P(t)$ at some future t. If

$$P(t, m)$$

was a linear sum

$$P(t, m) = \alpha f(t) + \beta m_t$$

then it must be that $\beta > 0$, which indicates that the momentum positively influences stock prices, making the equation consistent with momentum investing theory.

A good guess for $P(t, m)$, the *predictor* would be

$$P(t_{k+1}, m) = \alpha P(t_k) + \beta V(t_k) + \gamma m_{t_k}$$

This equation simply says that the current price is a linear sum of the previous price, the previous trading volume, and the momentum.

In the real world, instead of a function $P(t)$, instead there are several data points that give the price P for each point in time t. Thus, its derivative, the value for m_t, cannot be solved analytically, but it can be solved numerically. Instead of finding the exact value of the derivative, it is estimated.

One way to estimate the derivative of a function is the *centered difference formula*

$$m_t = \frac{\mathrm{d}P(t)}{\mathrm{d}t} = \frac{P(t+h) - P(t-h)}{2h}$$

This equation uses two data points before and after the point of interest to obtain an approximation of m_t.

With this equation, it is now possible to obtain the momentum without knowledge of $P(t)$, as long as $P(t+h)$ and $P(t-h)$ are known.

On account of the inaccuracy incurred by solving for the derivative analytically, the predictor is modified:

$$P(t_{k+1}, m) = \alpha P(t_k) + \beta V(t_k) + \gamma m_{t_k}^{h=22} + \sigma m_{t_k}^{h=7}$$

Now the predictor uses two momentums computed over different time intervals. Finding the value for these coefficients is as simple as applying linear least squares regression over a set of data points.

The group will find the optimal coefficients for seven companies in the Philippine stock exchange:

- AP
- CEB
- CHIB
- GTCAP
- MBT
- MEG
- RLC

3 Methodology

A dataset of the Philippine stock exchange was retrieved from Kaggle. The seven companies were chosen because of their relative volatility and their influence to the stock market as a whole.

Figure 1: Closing price graph of GTCAP, showing its volatility

3.1 Preprocessing

3.1.1 Removing irrelevant companies

There were many companies in the dataset. The rows containing companies irrelevant to the study were removed. The remaining rows were grouped by company, then separated into seven `.csv` files.

3.1.2 Finding the right price

The dataset did not have data for all the points of interest. For example, during holy week, there were no entries because the stock market was closed. If the group needed an entry at time t, the entry at the greatest t_{prev} was considered, where $t_{prev} > t$

3.1.3 Extracting relevant features

Just as there were too many rows, there were some irrelevant columns in the dataset. The group only needed the company name, the closing price, the trading volume, and the date at which the trade occurred. Only those columns (features) were considered for the rest of the research.

3.2 Graphing

Four python scripts located in `python_scripts/` were made to graph the four terms in the predictor over time. The module `matplotlib` was used to plot the resulting graphs.

3.3 Finding the right coefficients

The coefficients for the predictor for each of the companies were found using linear least squares regression.

4 Results and Discussion

Figure 2: Graph of closing prices

As shown in Figure 2, most of the companies have an unpredictable looking price graph.

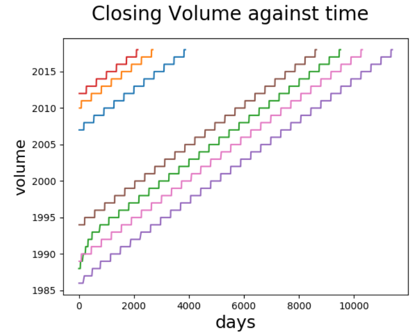

Figure 3: Graph of closing volumes

In Figure 3, it shows the total volume of trades the company went over time. The graph exhibits a "stairs" pattern, showing the spikes in buying and selling when prices fluctuate.

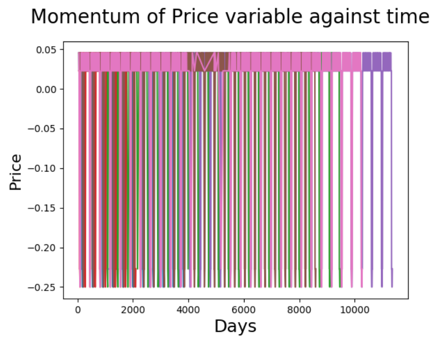

Figure 4: Graph of closing price momentum with $h = 22$

Stock prices tend to fluctuate the same way a heart beat wave fluctuates in an ECG monitor. The momentum graphs (shown in Figure 4 and Figure 5), which is a derivative of the price graph, looks like a bunch of candle sticks. Each abrupt spike/drop in price gets converted into an impulse in the momentum that shoots upward/downward.

It is worth noting that the momentum reaches farther into the negative side than the positive side, for both graphs. The momentum graph for $h = 22$ is less erratic and shows more stability, however the momentum graph for $h = 7$ reflects the instantaneous derivative better. Going to a larger stepsize means trading off accuracy for stability.

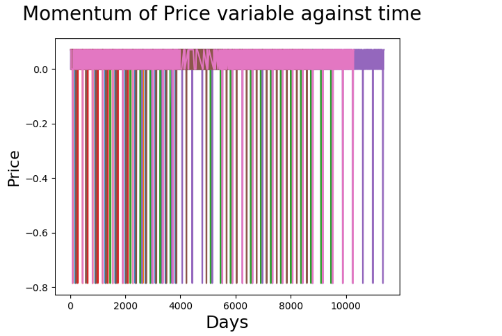

Figure 5: Graph of closing price momentum with $h = 72$

The table of optimal coefficients for the predictor

$$P(t_{k+1}, m) = \alpha P(t_k) + \beta V(t_k) + \gamma m_{t_k}^{h=22} + \sigma m_{t_k}^{h=7}$$

is shown in Figure 6.

For most of the companies, 7-day momentum has more influence than the 22-day momentum, not just because of it having a more positive coefficient, but also because the values for 7-day momentum are generally more extreme than 22-day momentum.

Cumulative closing trading volume has almost a zero coefficient on the predictor equation for all of the companies.

Previous price coefficient is almost one for all companies.

On average, the companies displayed similar tendencies except on the 22-day momentum coefficient, where they varied wildly.

5 Conclusion

7-day momentum has a bigger effect on future stock price than 22-day momentum. Cumulative closing trading volume has no effect on the future stock price. Future prices tend to stay close to the previous price, which indicates an overall stability in price in the stock of the surveyed companies.

References

[1] Heath M.,
Scientific Computing - An Introductory Survey

[2] Low, R.K.Y.; Tan, E. (2016). ”The Role of Analysts’ Forecasts in the Momentum Effect” . International Review of Financial Analysis.

[3] Philippine Stock Exchange Data,
https://www.kaggle.com/ianchute/philippine-stock-exchange-data

Company	α	β	σ	γ
AP	0.999909308676	0.000000000893	-0.002057696004	1.223340993478
CEB	1.223340993478	0.000000014819	0.040550002555	1.378957068841
CHIB	0.998817540958	0.000000000015	0.134914332815	1.571397313188
GTCAP	0.999805323406	0.000000601056	0.430392355966	1.158406101105
MBT	0.999110807743	0.000000017390	0.089714560749	1.419385112693
MEG	0.999526525917	0.000000000024	-0.000000000024	1.501996489010
RLC	0.999311080033	0.000000003205	0.115358802114	1.392324898465
Average	0.999463266205	0.000000091058	0.110422779598	1.378286853826

Figure 6: Table of computed optimal coefficient values