Tech Sector Stock Returns

IEOR 4150

Liuhaoyue Li (ll3351), Jiawen Huang (jh4179), Yuchen Fei (yf2515)

Abstract

In this project, we chose ten stocks in the technology sector which are considered as the top tech stocks primed to outperform over the next couple of years. We wished to gain some investment insights of this field through a statistical analysis of the ten stocks. With the analysis, we conducted a concise web app for better understanding of our clients. Firstly, we observed and analyzed the log-returns of each individual stock by displaying their histograms, normal probability plot, creating confidence intervals and performing regression analysis for them. After that, we set paired tests of every two stocks and performed regression on each of them to obtain their correlation.

Data Set

Our data set includes the daily open prices of every trading day of the ten selected NASDAQ technology companies from January 04, 2016 to November 11, 2019. A total of 985 data points collected for each stock were taken from Yahoo Finance. The stocks and the companies represented in our data set are:

Alphabet (GOOG, GOOGL)
Nvidia (NVDA)
Amazon (AMZN)
Rapid7 (RPD)
Splunk (SPLK)
PayPal (PYPL)
Apple (AAPL)
Synopsys (SNPS)
Micron (MU)
Microsoft (MSFT)

The above securities were chosen from a pool of NASDAQ tech stocks which look bullish with a “strong buy” analyst consensus. Stocks in the sectors have attracted the attention of investors recently.

Project Goals

The tech sector actually experienced some extremely tough times since the beginning of 2018. Even Google, the big guy, has struggled a lot during that period. The main concerns might consist of the lofty valuations of some tech stocks, discontinuous company growth, limited innovation, cyclical technology leading to limited development, events similar to dotcom bubble, government's policy, regulatory risk and etc.

However, if we focus on the longer-term outlook rather than the troubles at the moment, a new view emerges. Taking a look of the S&P 500 which is based on market cap of companies listed on NASDAQ or NYSE, the index has been tech heavy. Moreover, despite many professional analysts claim that the prices will keep low for a while, the biggest names in the sector are contributing most of the growth and leading to prospects that there might be a bright market in the not too far future. Therefore, people always believe that investing big name company can make profit in the long run. Additionally, as for the industrial internal influence, the big-name tech companies will affect each other in the stock market.

In this way, to assist showing solid reasons for investment decision making, we investigated a series of statistical questions to check if the statement is right by considering from two aspects: the behavior of the log-returns of single stock and the relationship between the log-returns of two stocks.

From the perspective of single stock, firstly we wanted to figure out if the histogram for our data by stock symbol presented an approximately normal distribution so that indirectly show that if the data were consistent with a random sample. To make it more visual, we also add a normal probability plot. After that, we wanted to find out the performance of the stock by calculating its expected return and the volatility of the stock measured by mean and variance of log-returns. To be more comprehensive, we should also compute the confidence interval for both mean and variance. Lastly, we were curious about if there any positive or negative in its log-returns and thus we performed a simple linear regression of the log-returns based on time.

From the perspective of every two stocks, the question coming first is that whether the means of the two are equal or not. The other question was whether the two stocks were correlated which can be achieved by performing a simple linear regression of one log-return on the other.

Web App Design

We used the Shiny Package in R to develop our web app. Our web app consisted of four parts: the introduction of our app; the overview output of the single stock analysis with manually inputted stock choice and confidence level; the overview output of the two-stock analysis with manually inputted stock choices and confidence level; as well as the "About Us" part for ending. Additionally, we also provided a drop-down menus for users to change the stock symbol, confidence level and so forth to view and compare the data more clearly.

Analysis

Single Stock Analyses

For checking whether the log-returns were normally distributed, the histogram was displayed. Even though the histogram resembled an approximately normal distribution, we still needed an additional probability density curve on the corresponding histogram to make it more convinced. Therefore, the point estimation of mean and variance of log-returns were required to prove the normal distribution. According to the results of the web app, both normal density curve and histogram fitted well except that the histograms have a fatter tail than the normal distribution. With these observations, we could still assume that the log-returns were approximately normal distributed. Also, with the Central Limit Theorem, the large number of the data size also helped us to hold our assumption.

Once the assumption of normality has been satisfied, we moved on to the statistical testing. The mean and variance could give us an insight of the data so that we could determine which stock has higher return and lower risk. The calculated means and variances of the log-returns of these stocks were shown in **Table1 Single Stocks Overview**. Based on these calculated mean and variances, we conducted the two-sided confidence interval for both mean and variance of each stock's log-returns (**Table1 Single Stocks Overview**). For this problem, both population mean and population variance were unknown, so we used t-distribution and Chi-distribution with n-1 degrees of freedom for assisting building the confidence interval. With $\bar{x}$ and s^2 denote the sample mean and sample variance respectively. The formulation of (1- α) % two-sided confidence for mean was $[\bar{x}-t_{\alpha/2,n-1}*s/\sqrt{n},\ \bar{x}+t_{\alpha/2,n-1}*s/\sqrt{n}]$, and that for variance was $[(n-1)*s^2/\chi^2, (n-1)*s^2/\chi^2]$. In this project, n = 985, α = 95%.

Table 1 Single Stocks Overview

Name	Mean	Var	Std	Mean_ci	Var_ci	Coef	Intercept	R_square
NVDA	0.000	0.000	0.022	[-0.0025 , 0.0029]	[0.0004 , 0.0005]	0.000	0.002	0.002
AAPL	0.001	0.000	0.011	[-0.0006 , 0.0021]	[0.0001 , 0.0002]	0.000	0.001	0.000
AMZN	0.000	0.000	0.014	[-0.0021 , 0.0014]	[0.0001 , 0.0002]	0.000	0.000	0.000
RPD	0.001	0.001	0.028	[-0.0027, 0.0041]	[0.0007 , 0.0008]	0.000	0.000	0.000
GOOG	0.000	0.000	0.012	[-0.0011 , 0.0017]	[0.0001 , 0.0002]	0.000	0.000	0.001
SPLK	0.000	0.000	0.021	[-0.0025 , 0.0027]	[0.0004 , 0.0005]	0.000	0.001	0.000
PYPL	0.000	0.000	0.015	[-0.0015 , 0.0020]	[0.0001 , 0.0002]	0.000	0.001	0.001
SNPS	0.000	0.000	0.012	[-0.0009 , 0.0019]	[0.0001 , 0.0002]	0.000	0.001	0.001
MU	0.000	0.001	0.023	[-0.0031 , 0.0025]	[0.0004 , 0.0005]	0.000	-0.001	0.000
MSFT	0.000	0.000	0.011	[-0.0012 , 0.0015]	[0.0001, 0.0002]	0.000	0.001	0.001

The 95% confidence interval for μ were small enough and showed that the population mean log-returns of all the stocks were around 0 with the upper and lower bounds not exceeding ±1 cent. Also, the 95% confidence interval for σ^2 were also very small, indicating the log-returns of all of the stocks had little variability and indirectly proved their low risks. According to the result, Microsoft (MSFT), Apple (AAPL) and Alphabet (Google) had lowest confidence interval for mean and variance while Nvidia (NVDA), Rapid7(RPD), Splunk

(SPLK) and MU had relatively larger confidence interval. The two clusters of stocks were the two extremes of risk.

We then performed a linear regression for each of the log-returns of stocks plotted against time (shown as **Table1 Single Stocks Overview**).

With the results of the table, the R^2 values for every linear regression was approximately zero which indicated that there did not appear to be a linear between each stock's log-returns and time. Also, referring to the scatter plots of the regression over time of each stock, the decentralization of these graphs also demonstrated the non-linearity. In this way, our consideration that utilizing time to predict the future log-returns might be useless and was a kind of "old wives tale".

Two Stock Analyses

Coming to the two stock analyses, as there were 362880 comparisons of each two stocks. Hereby, we selected two stocks, Nvidia (NVDA) and Microsoft (MSFT), one was in the cluster of lower risk while the other was in the cluster of higher, to conduct hypothesis testing and to build linear regressions respectively.

When comparing two stocks at a time, Nvidia (NVDA) and the other 9 stocks, we used a confidence level of 95% and the two-sample t-test with mean, variance unknown. And we assumed their means were equal in null hypothesis and unequal in alternative hypothesis to determine whether the stock returns were equivalent (**Table 2 Two Stocks Equality Test for NVDA**). After performing a series of tests on Nvidia and other nine stocks, we surprisingly found that we failed to reject our hypothesis that the stocks had the equal log returns. Thus we would assume that if the clients already chose Nvidia to invest, then the effect of choosing other nine stock was almost the same. Therefore, if the clients wanted to invest in the tech sector, he could choose these stocks randomly.

Table 2 Two Stocks Equality Test for NVDA

Name	Equality_Test_P_Value
AAPL	0.375
AMZN	0.364
RPD	0.596
GOOG	0.820
SPLK	0.929
PYPL	0.953
SNPS	0.639
MU	0.494
MSFT	0.892

Then, through building linear regression between Microsoft (MSFT) and the other nine, we could investigate in there any correlation between the log-returns of two stocks (**Table 3 Two Stocks Regression for MSFT**). Our null hypothesis is that coefficient between two stocks was zero, and the alternative hypothesis was not zero. After the series of tests performed, It is very clear that relationship between the two stocks is very evident as all of the coefficients between each pair are very closed to 1. And even the Synopsys and Microsoft had a relatively lower coefficient, the P-value of the test for them was really low such that we were able to reject the null hypothesis. It was very amazing that according to the results of the table, we had to reject each null hypothesis with 95% confidence level and the only problem was that the curve fitting is not that perfect as the R square of all pairs were never over 0.6. But to some degree, some of the observed variation can be explained by the model's inputs as the coefficient between two stocks was not zero. In a word, all of these data did prove that there might be some connection between two tech stocks randomly selected out of these ten stocks. Thus, in the future, investor can make use of one stock's return information to bring some insight about the other stock's return. For example, the growing trend of return of Nvidia may also imply the growing trend of that of Microsoft.

Table 3 Two Stocks Regression for MSFT

Name	Coef	P_value	R_square
NVDA	1.065	0.000	0.299
AAPL	0.627	0.000	0.400
AMZN	0.895	0.000	0.491
RPD	0.962	0.000	0.153
GOOG	0.759	0.000	0.527
SPLK	0.974	0.000	0.262
PYPL	0.803	0.000	0.375
SNPS	0.648	0.000	0.374
MU	0.920	0.000	0.202

Conclusion

After conducting our tests on the data, we could first draw the conclusion that the log returns of ten stocks we chose were consistent with a random sample, as it followed normal distribution. Based on this result, we could do further hypotheses testing.

As for analyzing the log-returns of single stock, we surprisingly found that the log return of all the ten stocks were close to zero, which meant that these tech stocks did not fit for long-run investment. This disproved our original assumption. We also conducted the regression between date and log return. Then we figured out that the R^2 of the regression was not close to zero, and there was no obvious pattern between them. Then we separated single stocks into high and low risk to conduct further testing.

By analyzing and comparing two stocks, we chose the lower risk representation Nvidia (NVDA) and higher risk representation Microsoft (MSFT) to conduct hypothesis testing and to build linear regressions respectively. First conclusion was that the difference of mean of each pairs is close to zero. In this way, we could not significantly prove that the other nine stocks had different log return with Nvidia. Nevertheless, after the simple linear regression, we found that the log return of Microsoft was relevant with the others. This point proved our assumption at the beginning.

To sum up, our testing results turned out to be mostly different with our initial guess. We finally found that in the three-year time period, invested the tech sector might not be a wise decision, and among these ten stocks, no matter which one you select, the log return would be almost the same. But investors can still make use of the information of one stock to predict the future trend of the other. That helps the analysis to be more comprehensive.

Reference

1. Banerjea, S., & Zacks. (n.d.). Why I'm Bullish on the Technology Sector. Retrieved from https://www.nasdaq.com/articles/why-im-bullish-technology-sector-2018-06-01.

2. Lefton, H., McCall, M., Jagerson, J., Hansen, W., McCall, M., Dec, & Dec. (2019, November 22). 10 Best Tech Stocks to Buy Now for 2025. Retrieved from https://investorplace.com/2019/11/10-tech-stocks-to-buy-now-for-2025/.