Association Between of Aspects of Quarterly Earnings Report and Short-Term Share Price for Large Market Capitalization US Companies

Jakob Danninger Jerry Xiao , Akansh Devendra

Abstract

Investing in the stock market as a beginner can be extremely difficult; there are so many variables to consider when investing, such as price, timing, other stocks, current events, and inflation. We decided to try to model a small slice of the stock market, by analyzing how select aspects of the top 50 (by market cap) US companies' earnings reports could affect change in stock price after quarterly earnings reports are released. We decided to analyze revenue, net income, gross profit, and earnings surprise, as they were the most widely available and relevant variables. Through modeling with K-nearest neighbors, we discovered that through using all four variables, we could predict how short-term price would change about 40% of the time, meaning the model was below our expectations. Through linear regression, we found that revenue, gross profit, and net income showed no linear correlation with short-term change in price, but earnings surprise showed a correlation with a 0.254140 coefficient of determination. Looking at Pearson's and Spearman's correlations, only earnings surprise showed statistically significant results, with Pearson correlation of 0.46812 with a p-value of 0.000607 and a Spearman correlation of 0.61269 with a p-value of 0.00000225 showing strong statistical evidence of an association since the p-value is less than our alpha of 0.05. Seeing some promise in earnings surprise, we analyzed earnings surprise vs. short-term change in price with a polynomial regression, which produced an r^2 value of 0.33933. Overall, there is very likely no correlation between revenue, net income, and gross profit with short-term change in stock price. However, earnings surprise has a limited effect on short term change in stock price as shown through the results of linear regression, Pearson's and Spearman's correlation, and polynomial regression, as earnings surprise performed far better than the other three variables. However, it needs further analysis, perhaps in conjunction with different aspects of the earnings report and external factors.

1. Introduction

The goal of investing in the stock market is to earn abnormal returns. [1] However this can prove quite difficult, especially when competing against other profit driven investors in a market that has yet to be fully understood. One such not fully understood aspect of the stock market is how the market reacts after an earnings report is released. This gap in knowledge has led countless investors to make poor decisions in relation to how they invest after a company's quarterly earnings are announced. Therefore we aim to analyze how different aspects of the quarterly earnings report along with investor expectations affect the change in short term stock price. Our hope is that through a better understanding of what affects share price changes after quarterly earnings, investors can make better informed decisions that will lead to a more stable, predictable and profitable market environment.

1.1 Data Sources

Data was collected primarily from 2 different sources: Yahoo Finance through the yfinance API, and Zacks Investment Research. Yahoo Finance provided data on the earnings report such as revenue, gross profit, and net income. Zacks provided information on surprise, which is how many percent did a company over/under perform compared to market predictions. Zacks is a reputable investment research firm that supplies data to many media companies such as Bloomberg, CapitalIQ, Morningstar, and The Wall Street Journal. [9] They are large, well respected, and established. Zacks is widely considered as the most important investment research firm and specializes in the US market, which is helpful since we only analyzed firms on US exchanges.

1.2 Approach

Initially, we wanted to create a machine learning model that can be used to predict how the market will react to quarterly earnings reports data. However, after making a K-Nearest Neighbor classifier, we found that its accuracy is questionable. Therefore we decided to analyze the data using polynomial and linear regression along with Pearson's and Spearman's correlation testing. The goal of linear and polynomial regression analysis was to analyze each different variable (revenue, gross profit, net income, and earnings surprise) in isolation, to see if there was any direct relationship. Pearson's and Spearman's correlations were used to find whether or not there was an association between the different variables. Only p-values below 0.05 were considered, as the rest could be regarded as statistically insignificant. In order to visualize our approach refer to figure 1, which diagrams the steps our data took from collection to analysis.

1.3 Summary and Insights

Our findings indicate that there is no correlation or association between change in net income, gross profit, or revenue and short term change in share price after earnings announcements. However, there is a limited correlation between surprise and short term change in share price. We also believe that there are other non earnings report factors that can lead to market swings, however these are hard to quantify and thus hard to analyze. Therefore, more research is needed, in order to be able to accurately predict market reactions to quarterly earnings reports.

2. Methods

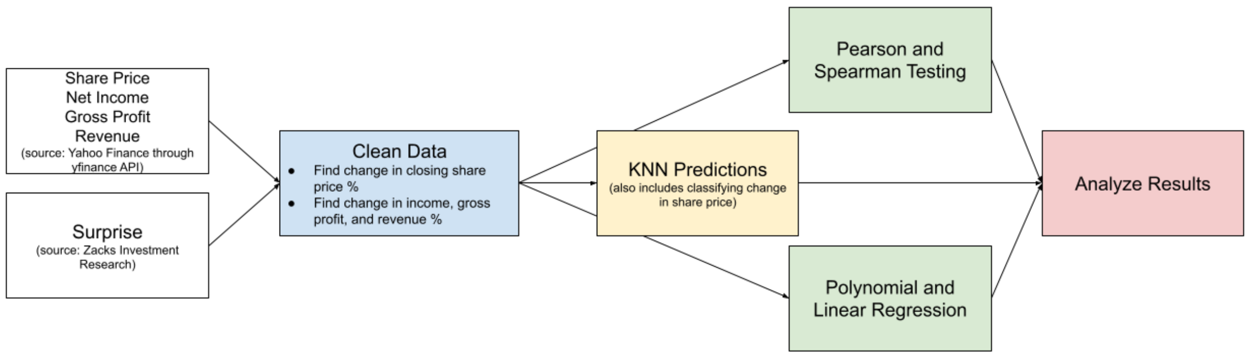

Figure 1: Project Workflow

2.1 Collection

In order to avoid issues with startups which tend to have strong valuations despite low sales, this report focuses on the top 50 US companies by market capitalization. Startup valuation is hard to predict and still not well understood. Their valuation is not based on current earnings, but rather future earnings. For instance, Nikola motors IPO valuation through a SPAC was 3.3 billion dollars, however Nikola had yet to start commercial production and has not delivered any vehicles. [11] From a data standpoint there are too many unquantifiable things in startup valuation such as the company's future plans and quality of leadership to warrant analyzing them.

The raw quarterly earnings data for these 50 companies was collected using the yfinance API which sources data from Yahoo Finance. The quarterly financials consists of 22 distinct features ranging from familiar financial information such as total revenue to obscure expenses such as accounting charges. Some companies were missing some features, therefore only the most widely available and influential features were chosen. We decided that gross profit, net income, and revenue were the most important features since they are generally the first thing investors look at. [10] The other features tend to be smaller after thoughts that are only important in special cases or specific industries.

Share price was also collected using the yfinance API and consists of 7 features, opening price, closing price, low, high, volume, dividends, and stock splits. We used the most recent closing price before the report was released and the most recent closing price the day after the report was released in order to give the market time to react to the report.

Surprise refers to how much an earnings report is above or below market predictions in terms of percentage. Surprise can vary depending on which source is used as a future earnings prediction. Surprise was collected from Zacks Investment Research and consists of 1 feature, the surprise when compared to Zacks earnings prediction. We chose Zacks for surprise data over other investment research firms since they are a market leader, supply financial data to media giants like Bloomberg, and they have a strong reputation.

All the relevant data was put into dataframes in order to be cleaned.

2.2 Data cleaning and Visualization

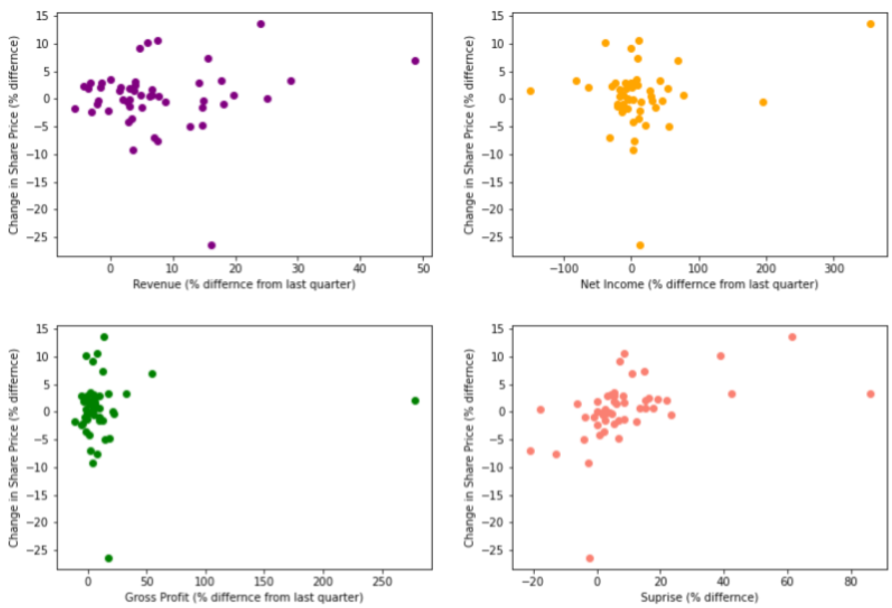

Figure 2: Scatter plots of each of the datasets created

We started by creating four different data sets, each corresponding to one of the four aspects of stock information that we chose to study (revenue, gross profit, net income and earnings surprise). For revenue, gross profit, and net income, the percent difference between the last two reported quarters was used as the independent variable. For earnings surprise, the data was already clean and remained unmodified. The dependent variable for each of the datasets was short-term change in stock price, which was calculated by finding the percentage change in closing share price between the trading day before and after the earnings report was released. This proved quite challenging since we had to adjust days for weekends and holidays (markets are closed then), but was solved by writing a custom function that used the datetime library.

Scatter plots for each of the data sets was then created, in order to get a general feel for the data. This is shown in figure 2 and will be discussed in further detail in the results section.

2.3 K-Nearest Neighbor

The K-Nearest Neighbor (KNN) model was created using the sklearn package in python. Since KNN only works to predict categorical data, the stock price change was split into 4 categories based on short-term percent change in price: less than -5%, -5% to 0%, 0% to 5%, and 5% or higher. A test size of 40% was utilized in this instance: out of all the sizes we tried, it tended to give the most consistent results. 1000 iterations of the KNN test were run, with randomized data in the test and train sets. The mean accuracy and 95% confidence interval of these iterations were then found.

2.4 Linear Regression

For each of the four datasets, we created a linear regression model with a 70-30 train-test split, using the sklearn package in python. 50 different trials were run (of which 5 were plotted) for each dataset, randomizing how the data was split among train and test each time. A table was then made for each dataset, containing the slope, mean-squared error, and coefficient of determination(r^2), for each of the 50 trials. The means of these three values were taken from each of the four tables.

2.5 Pearson's and Spearman's Correlations

For each dataset, the pearson and spearman correlations were found using the scipy.stats package in python. These results were put into a table.

2.6 Polynomial Regression

A polynomial regression fit was done on the earnings surprise data using the sklearn package in python. We only chose the earnings surprise dataset because all of the other datasets were extremely unlikely to have a meaningful polynomial fit (based on other tests). The graph and coefficient of determination were recorded.

3. Results

Figure 2 demonstrates the general shape of the datasets collected. The revenue and earnings surprise sets seem to suggest the possibility of linear relationships, and the other two graphs could suggest some sort of association. Based on these results, we decided to use a K-Nearest Neighbor classifier to combine these weaker relationships into a more accurate prediction model.

3.1 K-Nearest Neighbor

As stated in methods to get a KNN classifier to work, stock price change was classified into 4 different categories:

- Greater than 5% growth
- 0% to 5% growth
- -5% to 0% growth
- Less than -5% growth

The KNN code was run 1000 times using a test size of 40% and the accuracy of the classifier was recorded. The mean accuracy of the 1000 iterations of the KNN analysis was 39.5%, with a standard deviation of 0.0918470. The 95% confidence interval of accuracy was found to be between 39% and 40.1%, whereas random guessing would have an accuracy of 25%.

This shows that the KNN classifier works to a small degree, however, it is nowhere near accurate enough to consider it a successful model, being only 14.5% better than guessing. However, this did still show some promise, so we decided to focus on analyzing the individual features of our data through linear regression; there was a chance that, for example, only one of the four features actually correlated with short-term change in stock price, which would explain the poor performance of the KNN model. Linear regression would also provide more insight into exactly how, if at all, each variable was correlated with price change, as the slope of the correlation would be revealed.

3.2 Linear Regression

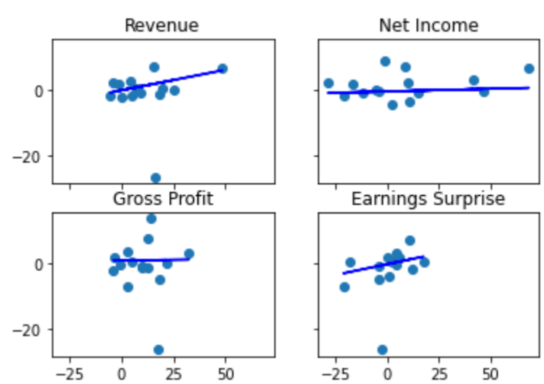

Figure 3a: Typical scatter and line as generated by linear regression models (x-axis: % change in respective variable, y-axis: % change in price)

	Avg. Over 50 Iterations		
	Coefficient	**Mean Squared Error**	**Coefficient of Determination (r^2)**
Change in Revenue vs. Short-Term Change in Price	0.021801	40.357974	-0.214951
Change in Gross Profit vs. Short-Term Change in Price	0.018145	41.736345	-0.338690
Change in Net Income vs. Short-Term Change in Price	-0.009973	30.907974	-0.177051
Change in Earnings Surprise vs. Short-Term Change in Price	0.236015	24.496766	0.254140

Figure 3b: Average values from linear regression models over 50 iterations of randomized 70-30 train-test split

Figure 3a represents typical graphs that would be generated by the linear regression models for all four datasets. From viewing the plots alone, it is unclear if linear regression is accurate or not. There seems to be a weak linear relationship, especially evident in the net income section of figure 3a. 50 iterations of a linear regression model were run for each dataset, and the coefficient, mean squared error, and coefficient of determination were collected for each iteration. We found that on average, there is a negative coefficient of determination for change for revenue, gross profit, and net income, vs. short-term change in price (-0.214951, -0.338690, and -0.177051 respectively). For change in surprise earnings, there is a positive coefficient of determination with change in price (0.254140). While this value shows a weak correlation, we deemed it worthwhile to further analyze it using Pearsons and Spearman correlations.

3.3 Pearson and Spearman Correlations

	Pearson's Correlation	Spearman's Correlation	Pearson Corr. P-value	Spearman Corr P-value
Change in Revenue vs. Short-Term Change in Price	0.1078590473697323	0.0830252100840336	0.45592878704849615	0.5664963036173906
Change in Gross Profit vs. Short-Term Change in Price	0.05840310895770906	0.12585834333733492	0.687042917439843	0.3838012075196312
Change in Net Income vs. Short-Term Change in Price	0.19226445126317204	0.06103241296518607	0.18101134198451307	0.67372460835961
Change in Earnings Surprise vs. Short-Term Change in Price	0.46812528529883146	0.6126917814918825	0.0006075023083251219	2.254033549682181e-06

Figure 4: Results of Pearson and Spearman correlations; significant results (pval < 0.05) are highlighted

Figure 4 shows the results of the Pearson and Spearman correlation tests. We only want to consider p-values that are less than 0.05, our alpha level, which would indicate that the test was statistically significant. Change in revenue, gross profit, and net income all have p-values greater than 0.05, which means they are not statistically significant, and can be disregarded. The only significant results are the change in earnings surprise vs. the short-term change in price, with a Pearson p-value of 0.000607 and correlation of 0.46812 as well as a Spearman p-value of 0.00000225 and correlation of 0.61269. This is the highest level of correlation that we've obtained so far. At a 0.61269 Spearman's correlation, change in earnings surprise vs. short term change in price is the most promising statistic in our tests. Again, while the correlation is not necessarily high, it is far higher than all of the other correlations found, warranting deeper study of earnings surprise through polynomial regression.

3.4 Polynomial Regression

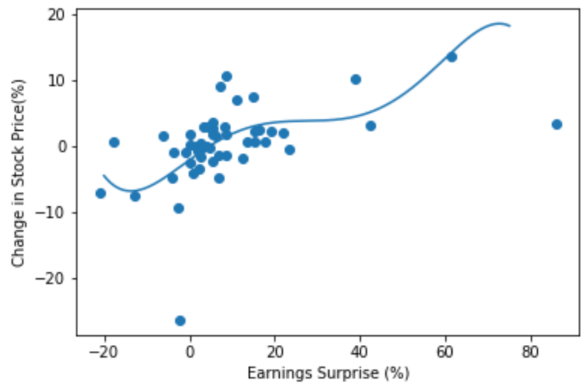

Figure 5: Polynomial regression with an r^ value of 0.33933

In order to study earnings surprise deeper, we decided to create a polynomial regression model on the entire dataset. While it seems like the regression fits the data quite well, the coefficient of determination was only 0.33933; still too low to draw any conclusions from. However, compared to the coefficients of determination seen in figure 3b, the r^2 value seen here is significantly higher. The regression could have had a higher r^2 value if the degree of the polynomial was increased, however that would lead to issues with overfitting and therefore we left it to 4.

4. Discussion

After looking through our results, it is clear that only surprise yields a statistically significant relationship with short term share price change due to having a p-value for the Pearson and Spearman test of less than our alpha is 0.05. This behavior of only surprise vs change in share price being statistically significant can be explained by the efficient market hypothesis where “market prices fully reflect all available information.” [8] This means that even before earnings are released, news around a company leads investors to predict earnings which is reflected in the share price. If a company meets expectations, their share price will not change since the prediction is priced in (already part of the share price). For example, if investors know the new Apple products are popular, they will predict Apple’s profits will go up a lot and that will increase the share price. Then when Apple’s earnings release and show profits have risen significantly, the market will not react since the higher profits were priced in when investors predicted there would be strong earnings. But, if the Apple earnings show only small or no profit growth, since the price previously rose due to optimistic predictions the share price will fall to reflect Apple’s true value. This explains why surprise, which is a measure of how much a company beats/missing expectations, has an association with share price change after earnings but no other features have an association.

However, even still this hypothesis is not fully accurate since surprise is only somewhat correlated to market reaction. Surprise is only a somewhat accurate predictor of stock performance after an earnings report for a few reasons.

Firstly, the surprise data we used is only an estimate. True surprise would entail understanding how millions of market participants predict earning will be, however this is an impossible task. Instead investment research firms like Zacks and Morningstar make predictions that many participants rely on and are in line with most investor's predictions. Hence, our surprise data is only somewhat accurate, hurting the accuracy of the KNN model.

Next, during earnings report calls with investors, many non-earnings report factors are announced and can affect how the market will react. For example, AAPL 2022 Q1 earnings showed a year over year increase in revenue, diluted earnings per share, and net income along with a surprise of 6.29%[1]. Judging by this one would expect the share price to rise, however prices fell 3.66% in response to Apple CFO Luca Maestri warned of future issues[2] with supply chain disruptions and low Chinese sales due to their lockdown measures.[3] Highly influential pieces of information like this are hard to collect and are not part of our model and therefore lowers our KNN classifier's accuracy.

Overall, the results point to the fact that out of the features we analyzed only surprise is important for predicting short term stock performance after quarterly earnings reports are published. This can be explained by the market pricing in their expectations thus the only thing that matters is how the quarterly earnings are relative to expectations. However, since there are so many other pieces of non-financial data that are discussed during quarterly earnings investor calls and it is impossible to get the true surprise value, surprise is only one of many factors that go into how markets react to earnings reports.

5. Conclusion

Through analyzing KNN, linear regression, Pearson's correlation, Spearman's correlation, and polynomial regression, we can not make any direct conclusions that change in revenue, change in net income, change in gross profit, or change in earnings surprise have any meaningful correlation with short-term change in stock price. However, despite still being low, earnings surprise shows far more correlation than the other three in all tests conducted (except KNN, which takes into account all 4 at once). While this still cannot lead to any conclusions, it could suggest that earnings surprise could be an incomplete factor- perhaps if analyzed alongside other factors we had not considered, it could create an accurate model. However, quantifying other variables could prove challenging, because the market responds to an incredible range of variables, such as current events, inflation, and politics. As such, it would be best to move forward by only studying quantifiable variables. It could also help to quantify some

---

[1] https://www.nasdaq.com/market-activity/stocks/aapl/earnings
[2] https://www.apple.com/investor/earnings-call/
[3] https://www.cnbc.com/2022/04/28/apple-aapl-earnings-q2-2022.html

variables- for example, perhaps the quality of operational performance could be made into a 10 point scale through expert analysis. With this, analyzing other aspects of the earnings report in conjunction with earnings surprise could be valuable, such as earnings per share, cash flow, or operational performance.[12]

Roles

Jakob: Data Collection, Data Cleaning, Discussion

Jerry: Methods, Results, Data Processing/Visualization

Akansh: Abstract, Introduction, Analysis

References

[1] Barone, A. (2021, April 25). *Abnormal Return*. Investopedia. Retrieved May 3, 2022, from https://www.investopedia.com/terms/a/abnormalreturn.asp

[2] Chen, J. (2022, February 8). *Earnings Surprise*. Investopedia. Retrieved May 3, 2022, from https://www.investopedia.com/terms/e/earningssurprise.asp

[3] Earnings announcement for AAPL. Nasdaq. (2022, April 28). Retrieved May 3, 2022, from https://www.nasdaq.com/market-activity/stocks/aapl/earnings

[4] Fink, J. (2021). A review of the Post-Earnings-Announcement Drift. *Journal of Behavioral and Experimental Finance*, *29*. https://doi.org/10.1016/j.jbef.2020.100446.

[5] Kifleswing, K. (2022, April 29). *Apple shares dip after company warns of a possible $8 billion hit from supply constraints*. CNBC. Retrieved May 3, 2022, from https://www.cnbc.com/2022/04/28/apple-aapl-earnings-q2-2022.html

[6] Likos, P. (2021, October 15). *How Earnings Affect Stock Prices*. U.S. News Money. Retrieved May 1, 2022, from https://money.usnews.com/investing/investing-101/articles/how-earnings-affect-stock-prices

[7] Likos, P. (2021, October 15). *How earnings affect stock prices*. Yahoo! News. Retrieved May 3, 2022, from https://news.yahoo.com/earnings-affect-stock-prices-191615204.html

[8] Lo, A. W. (2007). Efficient Markets hypothesis. *The New Palgrave Dictionary of Economics, 2012 Version*. https://doi.org/10.1057/9781137336583.0483

[9] Nasdaq. (n.d.). *About Zacks Investment Research*. data.nasdaq.com. Retrieved May 3, 2022, from https://data.nasdaq.com/publishers/ZACKS

[10] Radcliffe, B. (2022, March 14). *How to Decode a Company's Earnings Reports*. Investopedia. Retrieved May 3, 2022, from https://www.investopedia.com/articles/fundamental-analysis/10/decoding-earnings-reports.asp

[11] Renaissance Capital Renaissance Capital. (2020, June 3). *Electric vehicle startup Nikola set to begin trading under symbol NKLA following SPAC merger approval*. Nasdaq. Retrieved May 3, 2022, from https://www.nasdaq.com/articles/electric-vehicle-startup-nikola-set-to-begin-trading-under-symbol-nkla-following-spac

[12] Rodriguez, J. (2021, September 14). *What are earnings reports & what to look for in the quarterly statements*. Money Crashers. Retrieved May 3, 2022, from https://www.moneycrashers.com/what-are-quarterly-earnings-report-statements/

[13] Shah, Ravi, "The Impact of Earnings Announcements on Stock Prices: An Event Study to Examine Information Uncertainty and the Presence of Post Earnings Announcement Drift" (2020). *CMC Senior Theses*. 2532. https://scholarship.claremont.edu/cmc_theses/2532

[14] StockEarnings,Inc. (2022, May 3). *Sundial Growers (SNDL) earnings date*. StockEarnings.com. Retrieved May 3, 2022, from https://stocksearning.com/stocks/SNDL/earnings-date