
ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: “Big” Data Analytics
Proposal (or Final Report)

I. Abstract:

The goal of this project was to determine if short term price of stocks within the NASDAQ Biotechnology ETF (IBB) is correlated to earnings surprise. A simple linear regression model was made that compared the percent earnings surprise to the percent change in price one week leading up to the earnings release, one day after the earnings release from one day before the earnings release, one week after the earnings release, and one week after the earnings release from one week before the earnings release. The data was then standardized to changes in the IBB to account for systemic stock price changes, and linear regression models for this data were also formed. Overall, the regression models that were formed showed at best a weak positive correlation between short term stock price and earnings surprise, with the strongest correlation being between earnings surprise and stock price (adjusted to IBB) from one week before earnings release to one week after earnings release with an R^2 value of 0.3328. All other linear regressions had lower R^2 values, with the other linear regression models for data adjusted to the IBB having R^2 values having slightly higher R^2 values than data unadjusted to the IBB.

II. Introduction:

The goal of this project was to determine if short term price of stocks within the NASDAQ Biotechnology ETF (IBB) is correlated to earnings surprise. Only the top thirty holdings of the IBB were considered. Earnings surprise is the difference between the actual earnings release of a company and the consensus earnings projection made by financial institutes for that company.

It was hypothesized that there would be a weak positive correlation between earnings surprise and change in stock price one week leading up to earnings release. This is because investors that are more confident of their own earnings projections would buy/sell these companies' stocks, algorithm traders that trade on “momentum” would buy/sell these companies' stocks, and to some degree there would be some insider trading.

It was hypothesized that there would be a weak positive correlation between earnings surprise and change in stock price one day after the earnings release from one day before the earnings release. This is because there would be many investors that would simply react to earnings data and would buy or sell reactively.

It was hypothesized that there would be a weak positive correlation between earnings surprise and change in stock price one week after the earnings release because of the same reason as above.

Finally, it was hypothesized that there would be a strong positive correlation between earnings surprise and change in stock price one week after the earnings release from one week before the earnings release. This is because this is linear regression model would measure relatively long-term price change in comparison to the other models, so outliers would be less pronounced. The reasoning for the correlation is the same as above.

III. Method:

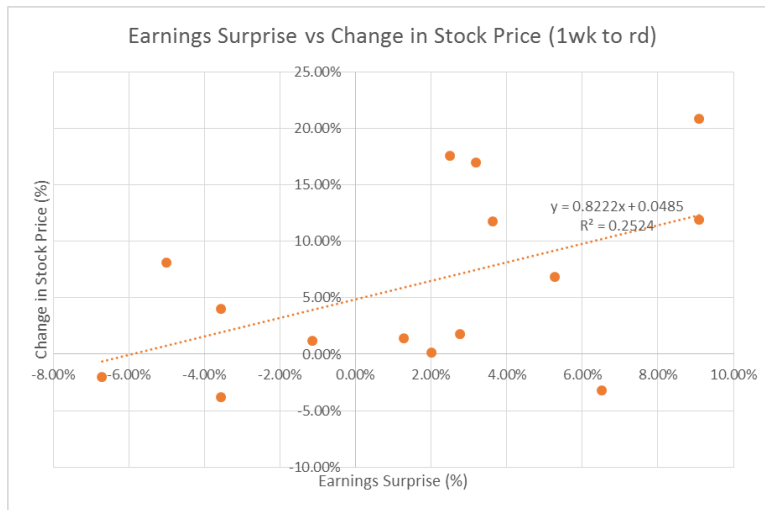
Earnings surprise data was collected from Zachs.com and historical stock prices for each company was collected from Yahoo! Finance.

Using Microsoft Excel, the percent change in stock price for each aforementioned time period was calculated and then plotted against earnings surprise. For each plot, a linear regression model was created. Initially, the data included the top 30 holdings of the IBB. However, there were some outliers, and it was decided that companies whose earnings surprise exceed a magnitude of 10% would be omitted from the linear regression model. This is because with such high/low earning surprise, the stock price would be very volatile in the short term.

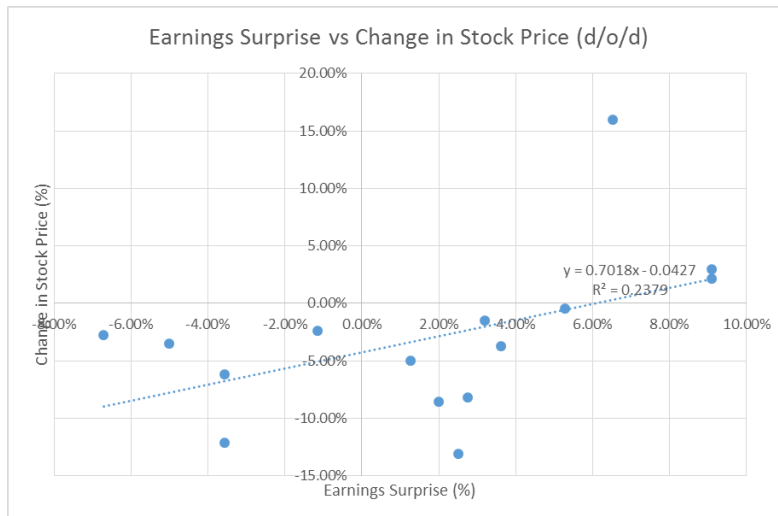
The data was then adjusted to changes in the IBB to account for systemic changes in stock price. This was done by calculating the change in the IBB during the same periods as the companies' stock prices and then subtracting the companies' price changes by the IBB's change. This is significant because companies within the same industry are often subject to the same macroeconomic trends or nonfinancial "shocks". For example, all companies in the IBB are affected when policy makers seriously discuss price gouging in the pharmaceutical industry or when the Federal Reserve announces changes in the interest rate. Linear regression models of the data adjusted to the IBB were made.

IV. Results

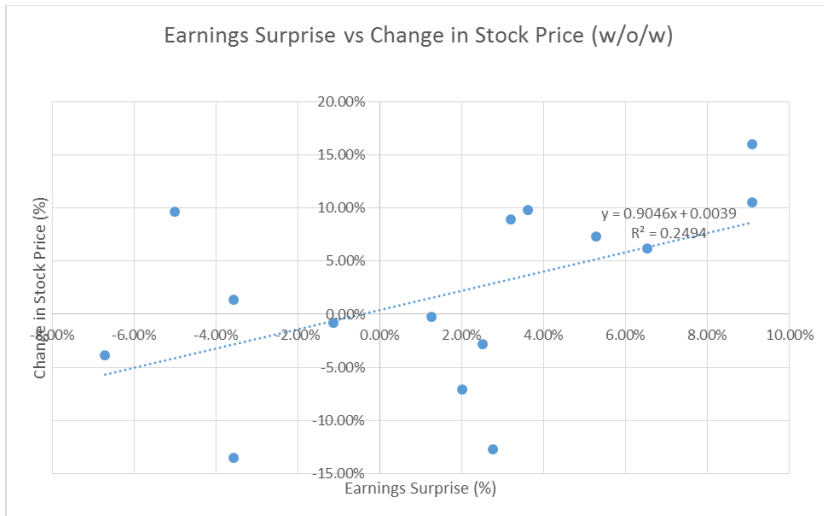
Data unadjusted to IBB without outliers:



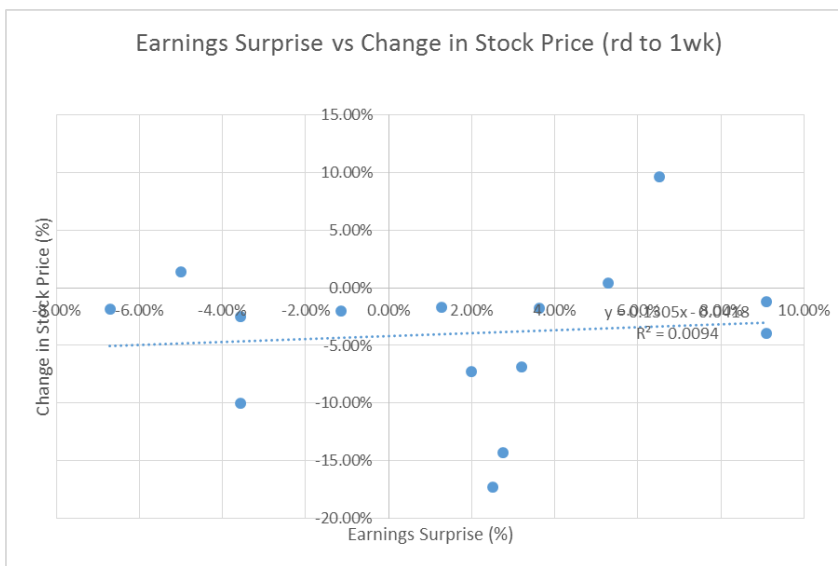
Earnings Surprise vs Change in Stock Price one week leading up to earnings release for data unadjusted to the IBB has a weak positive correlation with an R^2 value of 0.2524.



Earnings Surprise vs Change in Stock Price one day before earnings release to one day after earnings release for data unadjusted to the IBB has a weak positive correlation with an R^2 value of 0.2379.

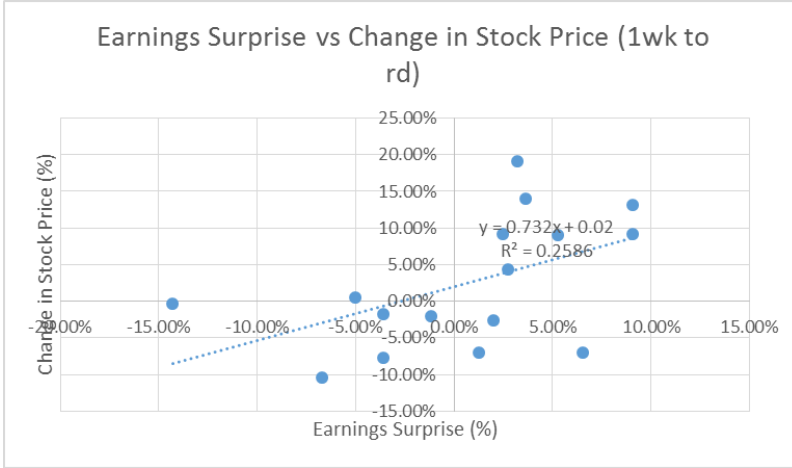


Earnings Surprise vs Change in Stock Price one week before earnings release to one week after earnings release for data unadjusted to the IBB has a weak positive correlation with an R^2 value of 0.2494.

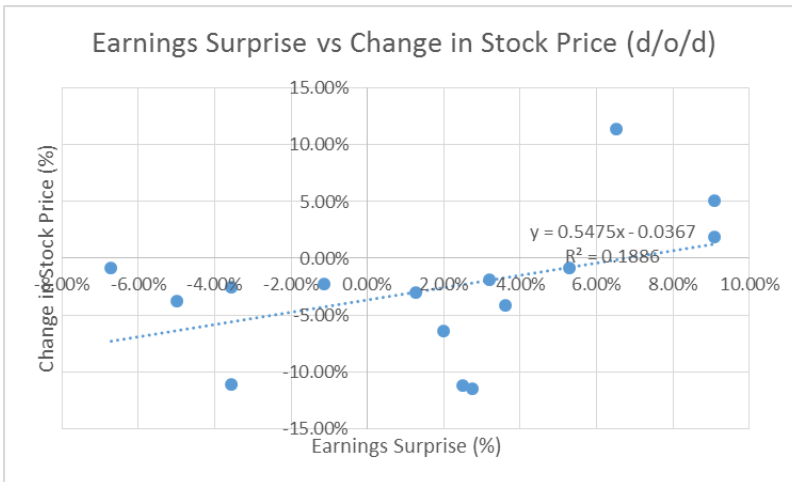


Earnings Surprise vs Change in Stock Price one week after earnings release for data unadjusted to the IBB has no correlation with an R^2 value of 0.0094.

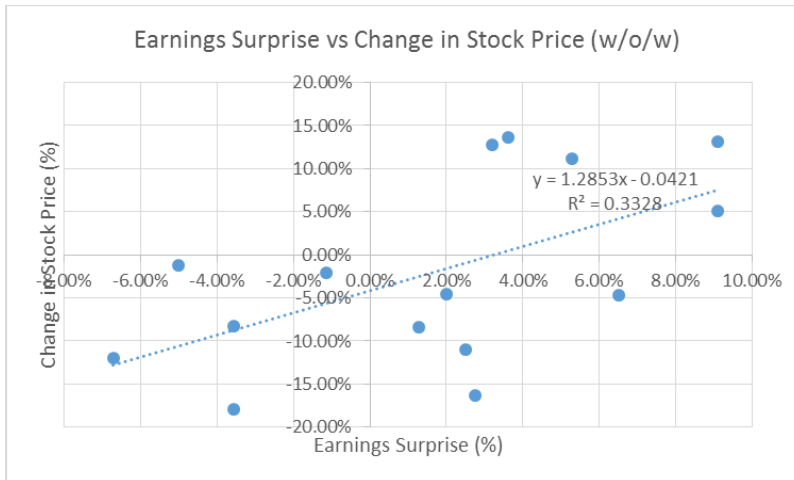
Data adjusted to IBB without outliers:



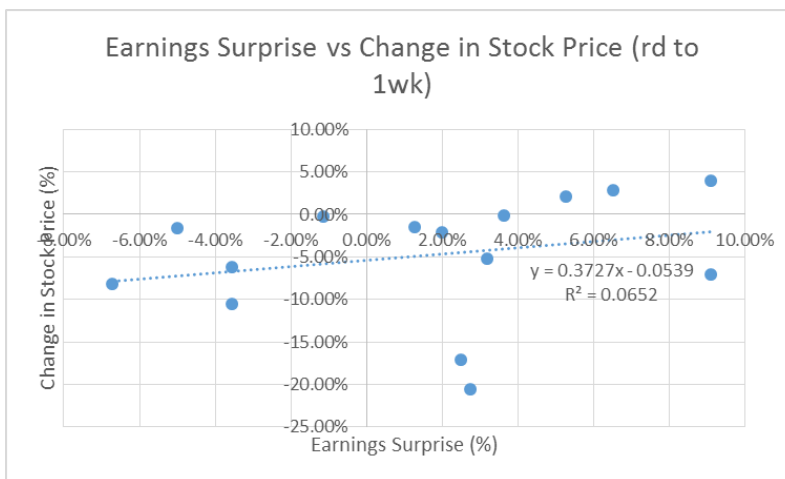
Earnings Surprise vs Change in Stock Price one week leading up to earnings release for data adjusted to the IBB has a weak positive correlation with an R^2 value of 0.2586.



Earnings Surprise vs Change in Stock Price one day before earnings release to one day after earnings release for data adjusted to the IBB has a weak positive correlation with an R^2 value of 0.1886.



Earnings Surprise vs Change in Stock Price one week before earnings release to one week after earnings release for data adjusted to the IBB has a weak positive correlation with an R^2 value of 0.3328.



Earnings Surprise vs Change in Stock Price one week after earnings release for data adjusted to the IBB has no correlation with an R^2 value of 0.0652.

V. Discussion:

Based on the data in the prior section, I accept the hypotheses that earnings surprise has a weak positive correlation with short term stock price one week leading up to earnings release and with stock price one day after earnings release from one day before earnings release.

I also reject the hypotheses that earnings surprise has a strong positive correlation with stock price one week before earnings release to one week after earnings

release and that earnings surprise has a weak positive correlation with stock price one week after earnings release.

There is however a weak positive correlation between earnings surprise and stock price change from one week before earnings release to one week after earnings release.

This data can be improved if there is a greater sample size. This is because the sample size that was fitted with linear regression after outliers were taken out was fairly small and the effect of outliers was very pronounced. I also believe that this data can be improved if one does a multiple linear regression model in which trade volume of the stock is also taken into account. This is significant because stocks that have higher trade volume are likely to be less volatile.

This data proves that the stock market is simply too complex is model with simple linear regression. There are too many factors that affect stock price to make a definitive model for stock price in the short term. It is not recommended to trade stocks based on simple models such as those made in this project.