Variance of Stock Returns

Gilbert Hall
Department of Physics, Boston University

Spring 2016

Abstract

Stock prices are often analyzed as a time series with the goal of predicting future prices. Data on the market indices are used to gauge the overall state of the market. I was interested in looking at the how individual stocks that make up such indices move since it is often just the overall index data that is quoted. I analyzed constituents of the S&P 100. I found the spread of the individual stock movement was correlated to several overall traits of the index.

1 Introduction

Stocks prices and market conditions are highly scrutinized because experts can use them to make money in the stock market. Often, the goal is to predict a stock price at a future point in time. The ability to do this accurately would be very valuable. The analysis used to make these predictions often involves analyzing stock prices as a time series. To get information on what the market is doing in general, analysts look at data on stock market indices. A stock market index is a collection of certain stocks meant to represent a particular sector of the market. A very commonly used index is the S&P 500, a collection of 500 large companies on the New York Stock Exchange or the NASDAQ Stock Market.

One of the most commonly analyzed pieces of data on the stock market is the returns or logarithmic returns of a stock or index. The return is a comparison between the price of a stock at the beginning and the end of a certain period of time, often one day. For a stock with price P_t on day t and P_{t-1} on the day before, the logarithmic return is $r_t = log(P_t/P_{t-1})$. The logarithmic return is zero if the stock price does not change, is positive if the price goes up, and negative if the price goes down. The time series volatility of an asset is the return squared, r_t^2. Since it is squared, it just gives an idea of how much an asset moved regardless of whether its value went up or down. A third measurement often applied to market indices is volume. This measures how many shares of something are traded in a certain period of time.

The goal of this project was to examine how the individual stocks in an index move; how closely they move to the average and whether the spread of the returns is correlated to any traits of the whole index. The stocks used are the constituents of the S&P 100, a commonly referenced stock market index. The dependent variable in this analysis was the standard deviation of the returns of the constituents of the S&P 100 on a certain day.

2 Hypothesis

The purpose of this analysis was to find correlation between the spread of the S&P 100 constituent returns and overall measures of the S&P 100. As mentioned, the dependent variable was the standard deviation of the constituent returns. The two most seemingly promising independent variables were the volatility and volume of the S&P 100 index. The planned method of analysis was to make two separate linear regression models, one for volatility and one for volume. The main question was whether or not the models would have statistically significant coefficients and if they do, would they be positive or negative.

The hypothesis for the model with S&P 100 volatility as the independent variable and return standard deviation as the dependent variable was that the model would have a positive coefficient for volatility; that is there would be a positive relationship between volatility and return standard deviation. The main reason for this hypothesis is that the volatility takes into account all the stocks in the index. Something could happen in a certain industry causing certain stocks to rise or fall while others are not affected. This would increase both the volatility and the return standard deviation. Also, there are many companies who are competitors or whose products can replace one another. As the stock of one of these companies went up, the others could go down, also increasing the volatility and return standard deviation.

The hypothesis for the model with S&P 100 volume as the independent variable and return standard deviation as the dependent variable was also that the model would have a positive coefficient; return standard deviation would increase with volume. The reason behind this is that the more trading that is going on, the more possible strategies the traders would have which cause a wider spread in the stock returns. However, it is possible that an increase in volume could signify an event that would cause a majority of stocks to go up or down. If this was the case, the standard deviation of returns might not have a positive correlation with volume.

3 Data

The data used for this analysis were daily measurements of the S&P 100 and its constituents. There is a wealth of data on the stock market available on the internet, so there was no problem finding the data. The data used was from trading days between March 2011 and March 2016, in total about 1,200 data points. First, the daily closing price of the S&P 100 was used to calculate the volatility. These data points were found on the S&P 100 website. The daily closing prices of the S&P 100 constituents were found on the same website. The first step to finding a variable measuring the spread of the returns was to calculate the daily return for each individual stock. Then, for each day the mean and standard deviation of the returns were calculated. The third piece of data, the S&P 100 volume, was not as clear-cut. The constituents of the S&P 100 are not all on the same exchange so there is not perfect, easy to find measure of the volume. The alternative used for this analysis was the volume of trading on the New York Stock Exchange. Many of the S&P 100 constituents are on the New York Stock Exchange, making it a good stand in for volume.

4 Results

The model created for this part of the analysis was a simple linear regression model with return standard deviation as the dependent variable and volume as the independent variable. The calculated parameter estimates for this model were 0.00568 for the intercept and 5.18×10^{-12} for the volume coefficient. On the face of it, 5.18×10^{-12} might seem too small to be significant, but it is reasonable because the volume is a relatively large number and the return standard deviation is relatively small. Also, the observed p-value is less than 0.0001, which means that although the parameter estimate is small, it is statistically significant. However, as show in Figure 1, the amount that the best-fit line rises over the span of the data is comparable to the variability for given values of volume. This means that although the model is statistically significant, it would not be too useful in predicting the return standard deviation for a given volume value.

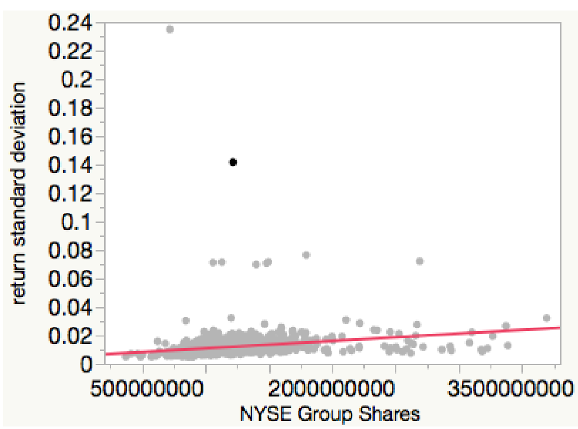

Figure 1: Volume vs. Return Standard Deviation

The same type of model was made for volatility as for volume; volatility was the independent variable and return standard deviation was the dependent variable. The calculated parameter estimates for this model were 0.0112 for the intercept and 6.38 for the volatility coefficient. This is estimate also has an observed p-value of less than 0.0001, meaning it is also statistically significant. Unfortunately, this model is similar to the previous one in that the amount that the best-fit line rises over the span of the data is comparable to the variability for given values of volatility, as shown in Figure 2. Like the first model, one probably could not use this model to accurately predict the standard deviation of the returns for a given volatility value.

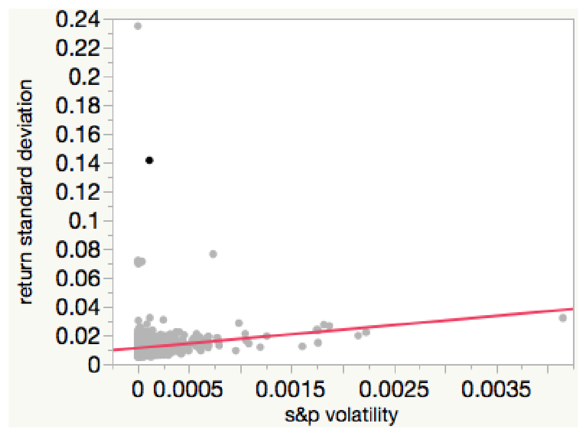

Figure 2: Volatility vs. Return Standard Deviation

5 Conclusion

The purpose of this analysis was to find correlation between the spread of the S&P100 component returns and other market characteristics. Two regression models were made, both with return standard deviation as the dependent variable, one with volume as the independent variable, and one with volatility as the independent variable. Both of these models show a positive relationship between the dependent and independent variables, meaning the standard deviation of the returns goes up as volume and volatility go up. Before the analysis, the hope was that an everyday trader would be able to use the results to help them make more informed decisions about the stock market. A trader would be able to look at certain aspects of the market and be able to predict how large the standard deviation of the returns of individual stocks would be. The use in this would be that if the spread of returns were greater, an individual stock would be less likely to follow predicted averages for indices and trading could be riskier or less predictable. Unfortunately, as mentioned in the results section, the model would not be very effective in predicting the exact return standard deviation for a given volume or volatility. For this reason, the results of this analysis would probably not be very useful to a trader who invests by hand. However, the models created were statistically significant so they might have some use. Trading with high-speed computers and algorithms is very common nowadays and is considered by some to be at the forefront of investing. The results from this analysis could be useful for traders using such methods. The algorithms used today take into account a multitude of market information to make decisions and these models could help make them more effective.