**Stock Data Analysis**

In this part, in order to do the hypothesis test that the volatility of weekly returns is same or not between the two industries, we have conducted F-test for checking if two samples’ variances are the same or they have equal volatilities. For calculating the F value for the hypothesis, we have calculated the mean, variance and sample size for every two samples. The result here is F= 1.067665135, it falls within C.V. This means we can’t reject the hypothesis, we consider the volatility equals between them.

In order to do the hypothesis test that two industries have the same mean of weekly returns, we have conducted T-test and Z-test for checking if the hypothesis is true. For calculating the T and Z value for the hypothesis, we have calculated the mean, variance and sample size. The result shows that the p values from T and Z are all larger than 0.01. This represents we accept the hypothesis, and we consider that they have the same mean.

In addition, we also conducted stock regression analysis on three different companies. The processes and outcomes are in the following:

1. **MSFT**

Here we have computed the weekly returns for the market index S&P 500 and the excess return(Ri-Rf). Then we build a regression model considering excess return as response variable and Rm-Rf as the factor. For MSFT, the result shows that the coefficient of Rm-Rf is 1.09. When it comes to checking the goodness of fit, Adjusted R square and Anova table both play critical roles. For MSFT, the result shows that the R square of fit is 0.543 and F is 123.44. And the coefficient of Rm-Rf is 1.09. Since the beta1 in every model is not equal to one, we can conclude that the Rm-Rf has impact on the volatilities of returns in MSFT.

Through the common tests for model assumptions, like residual diagnosis, line fit plot and normal probability plot, we can finally find out if our model follows the hypothesis. We can see from the plot results that the residual values of MSFT randomly scattered in the plot which we think they are normal distributed. The line fit plot and the normal probability plot also perform well in the graphs. These prove that the model of MSFT follows the regression assumption.

1. **IBM**

For IBM, the recall model shows that the coefficient of Rm-Rf is 1.121, which represents this index has impact on model.The recall model shows in IBM that R square is 0.398, F is 68.993. We can consider that the model of IBM doesn’t fit so well since R square is less than 0.5 and ANOVA table represents the index is important.

For IBM, the coefficient of Rm-Rf is 1.121. Since the beta1 in IBM model is not equal to one, we can conclude that the Rm-Rf has impact on the volatilities of returns for IBM. Then using the same measure for fit, IBM also performs well according to the three plots. Its plots obviously tell us their model can’t reject the assumption.

1. **T**

For company T, the final coefficient beta matches Rm-Rf is 0.678, which is smaller than the others. This means the index is not as important in the regression model for T as the other two stocks.When it comes to T, R square equals 0.152 and F is 19.434. T has the R square which is much smaller than 0.5, its model doesn’t fit as well as MSFT. But the result in Anova shows that the model is reasonable as well. When we move to company T, beta matches Rm-Rf is 0.678, which is not equal to one. That’s the reason why we need to consider the Rm-Rf in the brand new market-factor models. Then using the same measure to plot the regression, company T also performs well according to the three plots. The points scatted in the residual plot and the normal graph seems follow the normal distribution. We are supposed to accept the assumption.

According to the results above, we can see that MSFT has the best fitted regression model for its R square and the coefficient of Rm-Rf outperform IBM and T. IBM would be placed between them considering the goodness and fitness of its model. Company T has the worst fitted model among them, but Rm-Rf still has impact on its excess return index to some degree.

What’s more, we still find out that Rm-Rf performs significantly only in the models of Google, Facebook, MSFT, IBM and T for their F value is larger and p-value is much smaller than 0.01. When doing the comparison among these five famous companies, we find out the coefficient of Rm-Rf represents increase along with higner stock price. Google has the highest coefficient, with Facebook closely following behind. IBM and MSFT also show a strong relationship with Rm-Rf, and T performs more weakly. And this also tells us, only using market model to do the analysis or stock prediction is not enough for it may ignore many critical factors like Rm-Rf in some company, especially for famous companies with high stock price.