William Brickowski

STAT 3400

Brian Zaharatos

5/03/19

*An Exploration into the Fama and French 3 Factor Model*

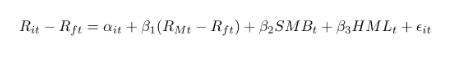
**Introduction**

The Fama and French Three-Factor Model is a model that was developed in the early 1990s and is used to price assets by predicting respective industry returns. It became broadly recognized after their publication of “Common Risk Factors in the Returns on Stocks and Bonds” in the *Journal of Financial Economic.* In 2013, Fama and French go on to receive a Nobel prize for their work in asset pricing.

Prior the Fama French model, academia in finance subscribed to the Capital Asset Pricing Model otherwise known as CAPM. The Capital Asset Pricing Model is a simple linear regression model with a single predictor. This predictor is equal to the returns of a portfolio that proxies for the market such as the S&P 500 minus the risk-free interest rate which is defined as the rate on a 1 month federal reserve treasury bills. The thesis behind the Capital Asset Pricing Model is that returns can only be generated through non-diversifiable risk which is also known as systematic or market risk. By regressing the returns in a particular industry against the returns of the market, the beta represents volatility with respect to the market. This form of risk is what drives returns.

Fama and French expand on the Capital Asset Pricing Model by including two other predictors which they believe also have a fundamental importance in understanding what drives returns in an industry. They added a size predictor and a value predictor because they showed, empirically, that small stocks tend to outperform large ones and value stocks tend to outperform growth stocks. According to Robert Brokamp, who quotes numbers from Ibbotson, between 1926 and 2017, “small-cap stocks returned 12.1%” whereas “large-cap stocks only returns 10.2%”. The difference in return between small and large cap stocks is known as the size premium which the size predictor (SMB) seeks to capture in order to more effectively describe market returns. According to Fidelity Investments, the “26-year annualized return of growth-oriented large-cap U.S. stock was 8.60 percent” and the “two value-oriented large-cap U.S. stock measures in this study, [the Russell 1000 Value Index and the Lipper US Index of Large Value funds], had an average return of 9.03 percent over the” same 26-year period between 1990 and 2015. The difference in returns between value stocks and growth stocks is known as the value premium. Fama and French include the value predictor (HML) in their model to capture the positive effect value has on generating returns.

The following is the Fama and French 3 Factor Model:



* Rit is the total return of a stock or portfolio, i at time t
* Rft is the risk free rate of return at time t
* RMt is the total market portfolio return at time tl
* Rit - Rft is expected excess return
* RMt - Rft is the excess return on the market portfolio (index)
* SMBt is the size premium (small minus big)
* HMLt is the value premium (high minus low)
* β1,2,3 refer to the factor coefficients. T

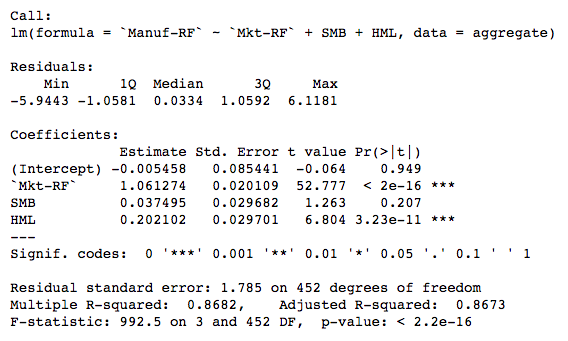
The three predictors of the Fama French model, Rm-Rf, SMB, and HML, are “constructed from size book-to-market benchmark portfolios that do not include hold ranges and do not incur transaction costs”. The excess returns on the market factor (Rm-Rf) is generated on the value-weighted return on all stocks traded on the NYSE, AMEX, and NASDAQ exchanges subtracted by the one-month Treasury rate. The size factor is calculated by taking the “average return on three small portfolios minus the average return on three big portfolios” : SMB = 1/3 (Small Value + Small Neutral + Small Growth) - 1/3 (Big Value + Big Neutral + Big Growth). The value factor is calculated by taking “the average return on two value portfolios minus the average return on two growth portfolios” : HML = 1/2 (Small Value + Big Value) - 1/2 (Small Growth + Big Growth).

In my research, I seek to investigate whether the multi-factor model with all three of the Fama French predictors is the best model to predict returns in the manufacturing, utility, and hitech industries. I am particularly interested in investigating if the Fama French model is the preferred model over the Capital Asset Pricing Model as well as if it does better in predicting industry returns. The final aspect of my analysis is to compare how well the preferred model does in predicting industry returns for data prior to 2013 as compared with afterwards. I chose to look at 2013 because this was the year that Fama and French received a Nobel prize for their multi-factor model. If prediction in the years after 2013 is better than the years preceding 2013, this may suggest that markets adapted to their work ex post facto.

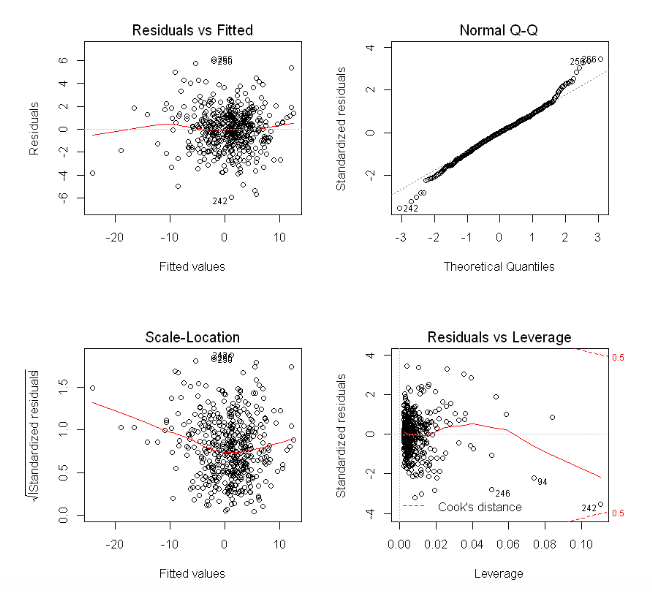
In order to carry out this project, I had to join a data set on monthly industry returns in the manufacturing, hitech, and utility sectors dating back to 1980 with a monthly predictor data set. I omitted the rows with na values. I then subtract the risk-free rate from all industry returns which will serve as the responses in my regression analysis.

**The Manufacturing Industry**

In order to select the preferred model, I carry out backward elimination, selection through minimization of AIC, and selection through maximization of Adjusted R-Squared. Backward elimination yields the model with the market returns predictor and the value predictor. The model selected through minimization of AIC returns the same model. The model selected through maximization of Adjusted R-Squared determines that the full model is the best. In order to come up with the best model between these two possibilities, I implement Leave-One-Out Cross Validation which returns a mean of the sum o the squared deviances. The function which minimizes this score, is the preferred model. The manufacturing industry model with all 3 factors has a Leave-One-Out Cross Validation score of 1486.684 (see Appendix I). The manufacturing industry model with Mkt-Rf and HML factors has a Leave-One-Out Cross Validation score of 24515.06 (see Appendix II). Hence, the preferred model is the full model with all three predictors.

****

Below are the diagnostic plots for the model with manufacturing return premiums as the response:

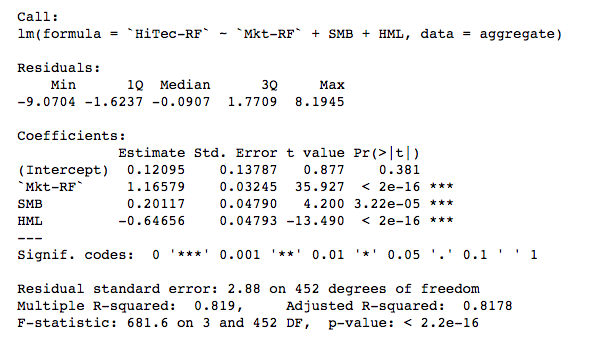


There does not appear to be any significant violations in the assumption of regression. There is no linear relationship or curvature in the residual vs fitted plot, hence the linearity assumption is not violated. There is also no trumpeting pattern in the residual vs fitted plot so the constant variance assumption is not violated. While there is some deviation in normality in the tails of the normal q-q plot, this is the least important of the other assumptions. All else equal, unless the sample size is small, results are reasonably robust to deviations in normality. The manufacturing industry model, has a condition number of 4.743328. Because the condition number is less than 30, there is no evidence of collinearity between predictors. Furthermore, there are no VIFs of greater than 10 which further suggests there is not evidence of collinearity. For this reason, no data transformations are necessary.

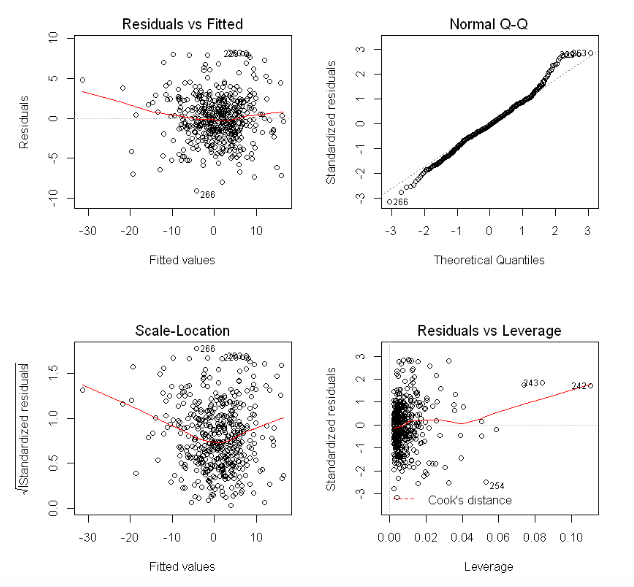
The Fama French 3 factor model has a Leave-One-Out Cross Validation score of 1486.684 whereas the Capital Asset Pricing Model has a Leave-One-Out Cross Validation score of 1609.472 (see Appendix III). This shows that the Fama French model is better than its predecessor, the Capital Asset Pricing Model. According to our preferred model, average returns in the manufacturing industry are -0.55%, if Mkt-Rf returns go up by 1.00% then manufacturing returns go up by 1.06%, if the SMB factor goes up by 1% then manufacturing returns go up by 0.037%, finally if the HML factor goes up by 1% then manufacturing returns go up by 0.20%. This model has an R-Squared of 86.82%.

**The Hitech Industry**

In this case, all three methods of selection yield the full model as the preferred model. For this reason, Leave-One-Out Cross Validation is not necessary to determine the best model. The full model is better at predicting hitech industry returns than the Capital Asset Pricing Model.



Below are the diagnostic plots for the model with hitech return premiums as the response:

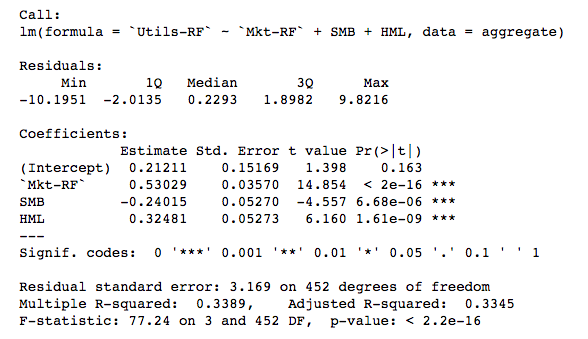


There does not appear to be any significant violations in the assumption of regression. There is no linear relationship or curvature in the residual vs fitted plot, hence the linearity assumption is not violated. On the residual vs fitted plot, there is some linear relationship in the range of the lower fitted values, however, this is a result of outliers pulling up the horizontal line and is hence not a legitimate issue. There is also no trumpeting pattern in the residual vs fitted plot so the constant variance assumption is not violated. While there is some deviation in normality in the tails of the normal q-q plot, this is the least important of the other assumptions. All else equal, unless the sample size is small, results are reasonably robust to deviations in normality. The hitech industry model, has a condition number of 4.743328. Because the condition number is less than 30, there is no evidence of collinearity between predictors. Furthermore, there are no VIFs of greater than 10 which further suggests there is not evidence of collinearity. For this reason, no data transformations are necessary.

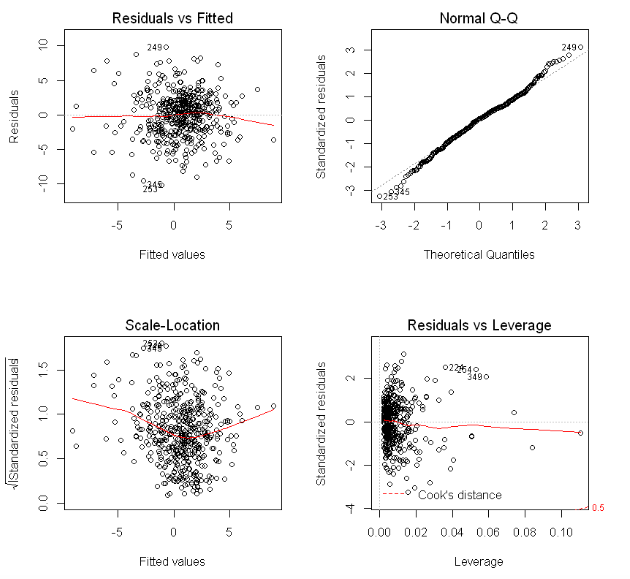
The hitech industry model with all three factors has a Leave-One-Out Cross Validation score of 3846.484 (see Appendix IV); whereas the hitech industry model with the Mkt-Rf factor only has a Leave-One-Out Cross Validation score of 5557.538 (see Appendix V). This shows that the Fama French model is better than its predecessor, the Capital Asset Pricing Model. According to our preferred model, average returns in the hitech industry are 12.09%, if Mkt-Rf returns go up by 1.00% then hitech returns go up by 1.17%, if the SMB factor goes up by 1% then hitech returns go up by 0.20%, finally if the HML factor goes up by 1% then hitech returns go down by 0.65%. This model has an R-Squared of 81.90%.

**The Utility Industry**

In this case, all three methods of selection yield the full model as the preferred model. For this reason, Leave-One-Out Cross Validation is not necessary to determine the best model. The full model is better at predicting hitech industry returns than the Capital Asset Pricing Model.



Below are the diagnostic plots for the model with utility return premiums as the response:



There does not appear to be any significant violations in the assumption of regression. There is no linear relationship or curvature in the residual vs fitted plot, hence the linearity assumption is not violated. There is also no trumpeting pattern in the residual vs fitted plot so the constant variance assumption is not violated. While there is some deviation in normality in the tails of the normal q-q plot, this is the least important of the other assumptions. All else equal, unless the sample size is small, results are reasonably robust to deviations in normality. The utility industry model, has a condition number of 4.743328. Because the condition number is less than 30, there is no evidence of collinearity between predictors. Furthermore, there are no VIFs of greater than 10 which further suggests there is not evidence of collinearity. For this reason, no data transformations are necessary.

The utility industry model with all three factors has a Leave-One-Out Cross Validation score of 4651.203 (see Appendix VI). The utility industry model with the Mkt-Rf factor only has a Leave-One-Out Cross Validation score of 5239.086 (see Appendix VII). This shows that the Fama French model is better than its predecessor, the Capital Asset Pricing Model. According to our preferred model, average returns in the utility industry are 21.21%, if Mkt-Rf returns go up by 1.00% then utility returns go up by 0.53%, if the SMB factor goes down up 1% then hitech returns go down by 0.24%, finally if the HML factor goes up by 1% then hitech returns go up by 0.32%. This model has an R-Squared of 33.89%.

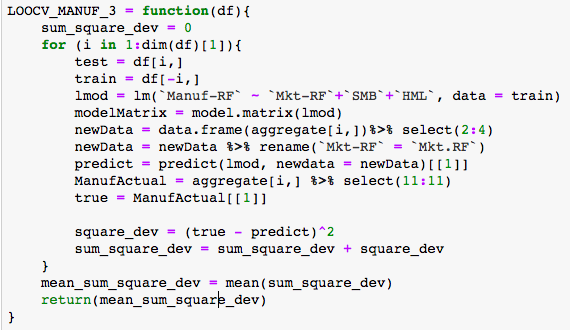
**Analysis into the Effect of Fama and French’s 2013 Nobel Prize on Prediction**

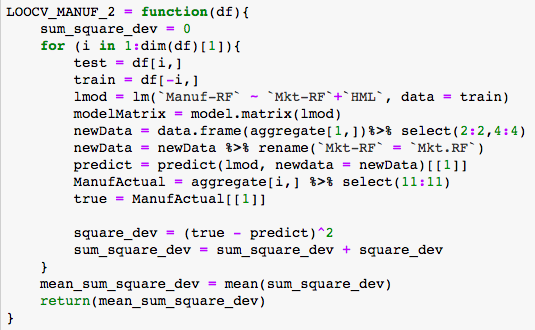
In order to analyze the effect that the 2013 Nobel Prize awarded to Fama and French on how well their model predicts returns in the manufacturing, hitech, and utility industry, I build two multi-factor models for each industry - one trained on pre-2013 data and one trained on post 2013 data. I then compare the R-Squared values and Mean Squared Error for the two models built in each respective industry. The utility industry model trained on pre-2013 data has a lower R-Squared than the utility industry model trained on post 2013 data meaning the model better explains returns in the utility industry after Fama and French had received the Nobel prize for their work. Additionally, the Mean Squared Error for the model that was trained and tested on pre-2013 data is higher than for the model that was trained and tested on post 2013 data. This means there is less prediction error after 2013, suggesting the 3-factor model is more accurate for predicting utility industry returns since Fama and French received their Nobel Prize. The hitech industry model also sees an increase in R-Squared going from pre to post 2013, however, Mean Squared Error rises so it is not conclusive as to whether or not the 3 factor model is better at making predictions in the hitech industry after the Nobel Prize had been awarded. This could potentially be a result of further predictors being needed to predict the often more volatile returns in the hitech industry. The manufacturing industry model actually gets worse after 2013 as it sees a decrease in R-Squared and an increase in Mean Squared Error. My thesis here is that in the years after the Nobel prize was awarded to Fama and French, many other models with more predictors were developed and so returns are most likely more accurately described by these such models. It may also be a coincidence that the returns in the manufacturing industry led to such results given that we were limited to very few years of data after 2013.

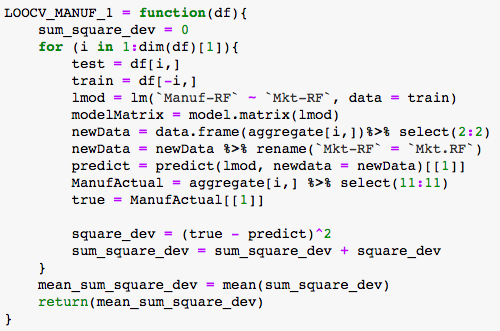
**Conclusion**

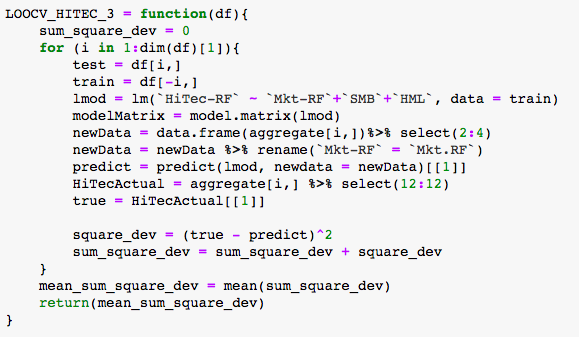
From this research, we have learned that the work of Fama and French is accurate in that their 3-factor model is more descriptive of market returns than its predecessor, the Capital Asset Pricing Model. We also have reason to believe that the winning of their 2013 Nobel Prize led to a self-fulfilling prophecy within the utility industry as the 3-factor model better explains returns in this industry after 2013 than it does prior. However, to confirm this, more analysis is needed. For example, breaking up the data into training and testing data in various ways and then retesting this hypothesis may be a valid choice. While the 3-factor model was the preferred model based on the three tested factors, it still had quite a low R-Squared value. I would be interested to see if adding further factors could aid in bringing this value up. Based on my analysis, it is still ambiguous as to whether model prediction got better or worse within for the hitech industry after 2013. However, I am interested utilizing other measures for prediction error other than MSE to base my ideas upon. It appears as if prediction based on the 3-factor model within manufacturing industry got worse after 2013. As previously stated, it would be interesting to break the data up into a training and testing set to determine if the model became better or worse after 2013. Another area that would be interesting to implement further research would have been to use K-Fold Cross Validation on top of Leave-One-Out Cross Validation to ensure that both forms of cross validation agreed upon the preferred model. Another limitation of this study is that there are not nearly as many years after 2013 as before which leaves room for outliers and statistical anomalies to impact results. Re-running this experiment in the future would mitigate such possible negative effects on the study.

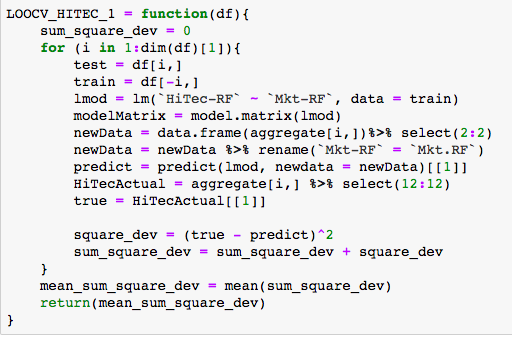
**Appendix**

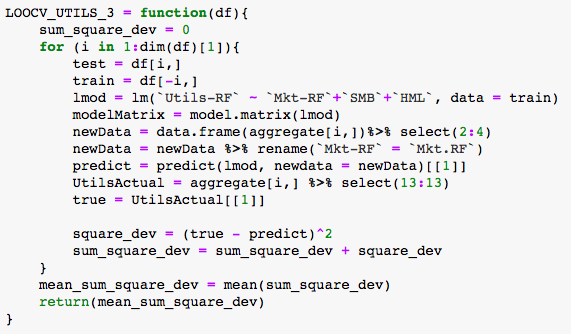
****

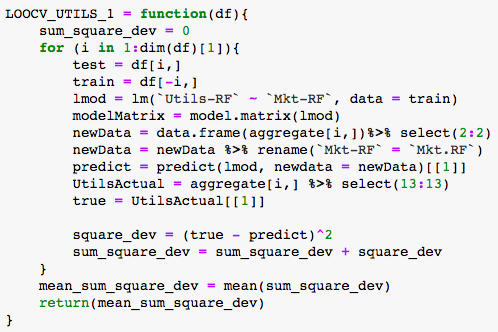
****

****

****

****

****

****

**Citations**

* Fama-French Three-Factor Model - Components, Formula & Uses. (n.d.). Retrieved from <https://corporatefinanceinstitute.com/resources/knowledge/finance/fama-french-three-factor-model/>
* Performance of Value versus Growth Stocks. (n.d.). Retrieved from <https://www.fidelity.com/learning-center/trading-investing/trading/value-investing-vs-growth-investing>
* Ross, S. (2019, April 30). What's the Difference Between Small Cap Stocks and Large Cap Stocks? Retrieved from <https://www.investopedia.com/articles/markets/022316/small-cap-vs-mid-cap-vs-large-cap-stocks-2016.asp>