

AI ETF fund analysis and Portfolio Management

**Dhivya Rao
Deep Borkar
Firas Obeid
Parth Patel**

**Dr. Liping Ma
FERM6V99: Methods in Financial Analytics
May 03, 2019**

Introduction:

Exchange-traded funds are a basket of securities such as stocks or bonds that tracks an underlying index with marginal tracking error not left out of the question. An exchange-traded fund is a marketable security meaning it can be bought and sold since the ETF has a price linked to it. ETFs can contain all types of investments including stocks, commodities, or bonds as stated earlier. These securities are more liquid than mutual funds and were designed to be traded more robustly than mutual funds. Unlike mutual funds, they are more liquid and as their name states are traded on exchanges. ETFs are attractive, nowadays, due to their low expense ratio for asset under which ranges between 0.01% - 1%.

Artificial Intelligence ETF

Artificial Intelligence is the specialization in computer science that aims at constructing intelligent machines and gadgets that work and react closely to humans. The ETF's that venture in AI focuses to gain from surged application and adoption of artificial intelligence. This comprises of artificial intelligence related to industrial or non-industrial robotics, automation and 3 Dimension printing, natural language processing, autonomous vehicles and social media.

AI ETF's are funds that fulfill at least one of the below mentioned scope:

1. They are funds that specifically invest in companies involved in the development of new products or services, technological improvements in scientific research related to artificial intelligence, or
2. They are funds that have at least 25 percent of their portfolio exposed to companies that shed huge spending on artificial intelligence research and (R&D) expense. For instance, on how companies like Amazon, Tesla Motors, Apple and Alphabet have huge funds invested on their product development.
3. They are funds that utilize artificial intelligence techniques to choose individual securities for concatenation into the fund.

We selected these ETF's based on their performance in terms of their Returns, Volatility and near average expense ratio.

ROBO GLOBAL ROBOTICS AND AUTOMATION INDEX ETF

Robo Fund Description:

ROBO fund captures a global index of corporations involved in robotics and automation. The ticker for this fund is ROBO which is primarily exchanged at NYSE ARCA. The fund came to inception on 10/21/2013. The fund has an expense ratio of 0.95%. The average daily share volume of ROBO constitutes 193,955 with an average daily collection of \$7.77 million.

Tracked Index: ROBO Global

Asset Allocation:

The fund has net assets worth \$1,499,578,289.40 with outstanding shares of 36,200,000. The present Net Asset value is \$41.42 as on 04/26/19. The fund has major investments in Industrial and technology sector. Yaskawa Electric corporation and OMRON corporation are their top two holders.

ISHARE S&P ASIA ETF

Ishare S&P Fund Description

The fund measures the performance and results of 50 high capitalized Asian corporations. The fund tracks the S&P Asia 50TM. This index is entirely float adjusted index that is weighted based on the market capitalization in order to estimate the performance of 50 top companies in four Asian countries regions namely Hongkong, Singapore, South Korea and Taiwan as derived by the S&P Dow Jones Indices LLC (the “Index Provider” or “SPDJI”)

Tracked Index: S&P Asia 50 Index

Asset Allocation:

The fund has net assets worth \$1.09 Billion with outstanding shares of 604,700,000. The present Net Asset value is \$62.70 as on 04/26/19. The fund has major investments in financial and technology sector. Tencent holdings Ltd and Samsung electrical co. Ltd are their top two holders.

QQQ INVESCO

QQQ Fund Description

Invesco QQQ ETF is based on Nasdaq 100 Index. The ticker of the fund is QQQ. The fund comprises of all the stocks in the above-mentioned index. The index has 100 largest international

and domestic corporations involved in the non-financial sector that is listed on the Nasdaq on the basis of its market segmentation. The fund came to inception on 10th March 1999. The fund has about 103 stocks from the non-financial sector. The fund has an expense ratio of 0.20%. The fund has total net asset worth \$74,002,284,009. The average daily share volume of QQQ Invesco constitutes 28,885,912 shares with an average daily collection of \$5.18 Billion

Tracked Index: NASDAQ 100 Index

Asset Allocation:

The fund has net assets worth \$74,002,284,009 with outstanding shares of 399.05MM. The present Net Asset value is \$190.45 as on 04/26/19. The fund has major investments in consumer cyclicals and technology sector. Microsoft corporation and Apple Inc. are their top two holder.

VGT ETF

VGT Fund Description

The investment seeks to trace the performance of a benchmark index. The fund employs an investment approach designed to trace the performance of the MSCI U.S.A. Investable Market Index/Information Technology 25/50

Tracked Index: MSCI U.S.A. Investable Market Index/Information Technology 25/50

Fund Allocation:

The fund has net assets worth 21.57 Billion with outstanding shares of 759,206,416. The present Net Asset value is \$212.95 as on 04/26/19. The fund has major investments in software and Information technology and Computer gadgets and technology sector. Microsoft corporation and Apple Inc. are their top two holder.

AIRR: First Trust RBA American Industrial Renaissance ETF

AIRR Fund Description:

The fund focuses on small- and medium sized capital industrial service and support companies and local banks in order to support them from a US industrial revival.

The issuer of the fund is First Trust. This open-ended fund came to inception on 03/10/14. The ticker of the fund is AIRR. The average daily trading share volume of AIRR constitutes 23,565 with an average daily collection of \$594.14K. The fund has an expense ratio of 0.70%.

Index Tracked: Richard Bernstein Advisors American Industrial Renaissance

Asset Allocation:

The fund has net assets worth \$79,762,067 with outstanding shares of 3,100,002. The present Net Asset value is \$25.73 as on as of 3/29/2019. . The fund has major investments in machinery & gadgets and Construction sector. Federal Signal corporation and EMCOR Group Inc. are their top two holders.

Descriptive Statistics:

For our analysis we choose 5 AI ETFs based on their performance relative to another traded ETF in the AI field. We perform our analysis on a cross – sectional and time-series data.

After extracting financial series data, adjusted closing prices of the stated ETFs and indices, and their respective log weekly returns are plotted, in addition to the two market indices (S&P 500 and Dow Jones technology index) and a distribution is fitted for each graph overlaid on a histogram where the bins are the weekly log returns. The 13-week treasury bill weekly rates are also extracted to derive the risk-free rate for our analysis. All tickers were imported using yahoo finance web page. However, we imported using CSV format five tickers, which are “SPASIA50”, “RBAAIR”, “DJUSTC”, “MSCI” & “ROBO.STOX”. The CSV files of the weekly return were downloaded from the Thompson Reuters platform (EIKON).

Indices outperform the ETFs that track them except for four out of the five cases, which is the AIRR, QQQ, ROBO and VGT where both ETFs beat the indexes they track. The respective annual return for AIRR, QQQ, ROBO and VGT is 4.85%, 15.63%, 7.34% and 17.33%.

However, the annual return for the tracked indexes are 2.85%, 15.28%, 6.96% and 15.52%. VGT has a higher return among all ETFs, and even market or tracked indices. The median of log

returns for all ETFs is above their respective mean except for the AIRR ETF. We observe negative skewness for these ETFs however, AIRR has a mean greater than its median, and that is evident in its 0.219 skewness. AIRR is an attractive ETF due to its positive skewness, not to mention also to the index it tracks since return distribution with positive skewness has frequent small losses and a few extreme gains.

The natural logarithm of returns shows a normal distribution among all ETFs and indices that were plotted. Most financial analysis assume of normally distributed returns and lognormal price distributions; however, we observe different levels of skewness and kurtosis among our findings that show a less accurate normal distribution among the returns that were plotted. Skewness is an important central moment measure that reveals the symmetry of the distribution of returns, in our case. QQQ, and VGT ETFs returns are moderately skewed since their skewness lies between $(-0.5, -1)$. QQQ's tracked index, NASDAQ-100, follows the same since it has a skewness of -0.776 , same is implied for MSCI and ROBO-STOX which are moderately negative skewed at -0.734 and -0.51 respectively. ROBO, AIA and AIRR have approximately symmetric return distribution for respective skewness of -0.49 , -0.31 and 0.21 . The S&P Asia 50 and RBAAIR follow the same. Moving on to the market indices, the S&P 500 and the Dow Jones technology index have a moderately symmetric distribution.

We find that from comparing central tendencies and dispersion of ETFs, all analyzed ETFs track and trail the central moments measures of the indices that they track. Moreover, comparing with the Dow Jones technology index, QQQ ETF best trails the latter as it tracks the Nasdaq-100 composite, which constitute 103 equities stock situated in the technology sector. VGT also trails the DJUSTC since it tracks the MSCI fund, which constitute all securities in the stated index that are classified in the Information Technology sector. Thus, it is evident that QQQ and VGT, with their mirroring in central tendencies and distribution to NASDAQ-100 and MSCI respectively, also track Dow Jones Tech Index.

Moving on to kurtosis that captures the fourth central moment, all ETFs and indices have a leptokurtic distribution due their excess kurtosis that exceeds that of a normal distribution. AIRR and its tracked index RBAAIR have the highest kurtosis among the analyzed funds/indices suggesting that they have infrequent extreme deviations.

The shapiro test is performed on the log returns to check if the respective ETF and the indices they track, follow a normal distribution. The p_values obtained show clear evidence that returns of the ETFs or their indices are not normally distributed since the null hypothesis of returns being normal was rejected in every security that was passed to the test. Moreover, the sharpe ratio for the ETFs and the indices are calculated. VGT has the highest annual sharpe ratio among all five funds and MSCI (the tracked index by MSCI) has the highest sharpe ratio among other tracked indices. The respective sharpe ratios of both are 0.998 and 0.927. VGT outperformed its index with a higher sharpe ratio making it an appealing investment in regard to the studied ratio measure.

Hypothesis Testing and Tracking Error:

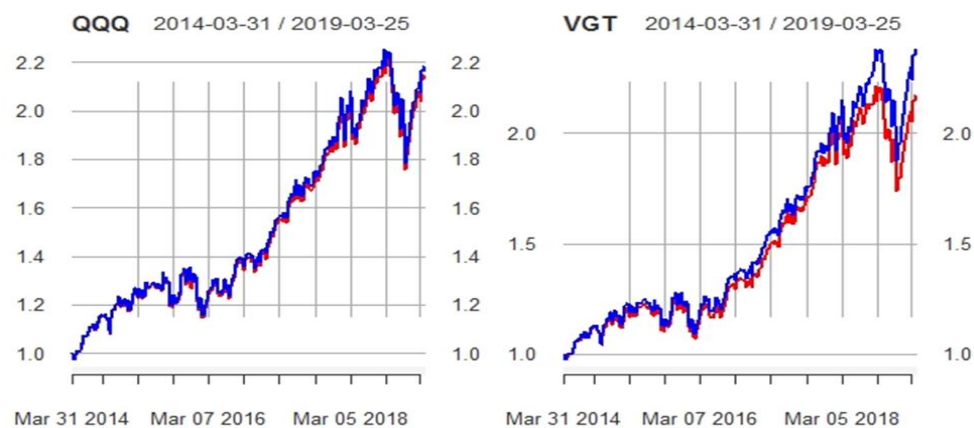
Nevertheless, a hypothesis test was generated to check if difference is zero between the weekly means off the ETFs and their respective tracked funds. Two data samples are matched if they come from repeated observations of the same subject. Thus, we assume that the data populations follow the normal distribution just for our hypothesis. We use the paired t-test, to obtain an interval estimate of the difference of the population means. Using the R programming language, the 5 ETFs and the 5 tracked indices are passed through a loop to perform the “Student’s-t test” to find the 95% confidence interval estimate of the difference between the stated weekly two sample means between "2014-04-01" and "2019-04-01". The test for all ETFs suggested that the mean between any ETF and its tracked index is greater or equal to zero since at a 5% level of significance, from the sample data, there is enough evidence to accept our null hypothesis ($H_0: \mu(\text{difference}) \geq 0$), which suggests that the means are greater than zero, in our case for a p_value > 0.05(alpha) in every case. The slight variations can be explained by the tracking error.

An investor requires a performance measure to decide which ETF he/she wants to invest in. Since an ETF tracks a specific index and tries to replicate its performance, they are evaluated relative to the index they track. Tracking index is a simple measure which is the difference between the performance of the ETF and the Index they track. It is calculated by taking the difference of the ETF return and the Index return over the same time period.

For the ETFs discussed above, the following is their performance analysis.



These are the cumulative returns calculated over 5 years for the ETF (Blue) and their Index (Red). As per the chart the AIA ETF the returns are mostly trailing its index, thus underperforming. Whereas for the AIRR ETF, both the fund and the index closely following a similar return. This indicates that AIRR is performing better compared to AIA.



In the above chart, the ETF QQQ is also following the index returns closely with its index. On the other hand, VGT ETF is outperforming its index which is an indicator of a good performance. The Tracking difference just tells us whether the fund is doing better than its index

for that period. It usually trails the index as an ETF charges a fee so if hypothetically the ETF followed the index perfectly, its returns would trail the index by the management fee.

Since markets are assumed to be efficient, a return higher than the market is considered abnormal return. Like in the case of VGT which is performing fairly better than the MSCI Information Technology Index.

Although tracking difference is useful, it does not give a complete picture of the fund's performance. Hence, the Tracking Error is used as a measure to assess the overall variability in performance of the fund versus the index it tracks. This measure is also indicative of the fund manager's performance as it is his/her duty to replicate a certain index efficiently and generate returns more than or at least equal to its index.

Tracking Error is the volatility of the difference between the ETF return and the index return.

$$TE = \sqrt{\frac{\sum_{i=1}^n (R_p - R_b)^2}{N-1}}$$

Where:

TE = Tracking Error

R_p = Return of Manager or Fund

R_b = Return of Benchmark

N = Number of Return Periods

These are the Tracking Errors for the ETF funds we selected:

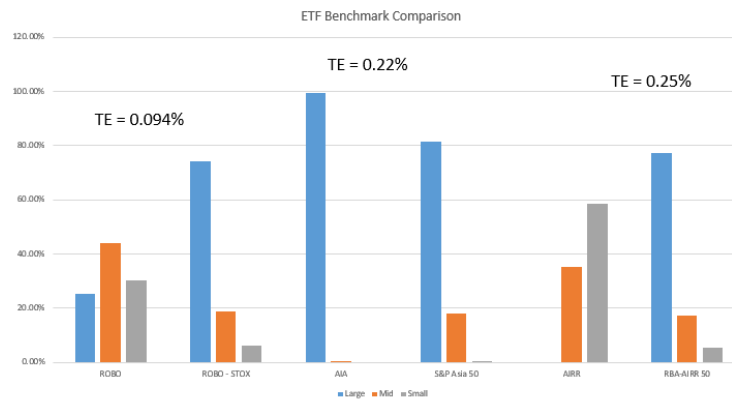
ETF	Tracking Index	Tracking Error (Weekly)
ROBO	ROBO – STOXX GLOBAL	0.094%
AIA	S&P Asia 50 Index	0.22%
AIRR	RBA-AIRR Index	0.25%
QQQ	Nasdaq 100 Index	0.017%
VGT	MSCI World IT Index	0.019%

The above table tells us that AIA and AIRR have high tracking error compared to the other funds. This means that their performance has been more volatile relative to the index they follow. QQQ and VGT has the least volatile performance versus its index and can make a good investment if they are able to generate returns higher than the index. There are several factors that affect the Tracking Error and can be used to improve the tracking error as well. The two main factors that affect the tracking error are Expense ratio and degree of index replication. One of the reasons ETFs are attractive investments are that they have a lower expense ratio than

mutual funds. If we look at the expense ratios for our 5 funds, we can observe that the funds with a high expense ratio have a relatively high Tracking Error as well.

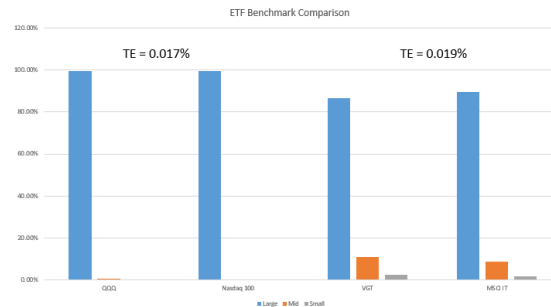
ETF	Tracking Index	Tracking Error (Weekly)	Expense Ratio
ROBO	ROBO – STOXX GLOBAL	0.094%	0.95%
AIA	S&P Asia 50 Index	0.22%	0.50%
AIRR	RBA-AIRR Index	0.25%	0.70%
QQQ	Nasdaq 100 Index	0.017%	0.20%
VGT	MSCI World IT Index	0.019%	0.10%

It can be observed that QQQ and VGT offer the lowest expense ratio and consequently have the lowest Tracking Error among the funds. Managers can reduce the expense ratio to control the tracking error and improve the performance of the fund. An ETF tries to replicate an index and tracks it over time. But it is almost impossible for the fund to invest exactly as the index as it can be too big of an investment and the manager while tracking majority of the index has another investment strategy. This can affect the funds tracking error significantly.



The above chart depicts the investment weights of ETF and their benchmark index in large cap, mid cap, and small cap stocks. Here, we can see that the ETFs ROBO, AIA and AIRR have a different approach in investing that their indexes. ROBO ETF invests more in Mid-cap and its index has more weightage to large-cap. AIA ETF invests majority of its funds in the large-cap more than its index. On the other hand, AIRR invests all its funds in Mid-cap and Small-cap. This significant difference of investment strategy affects the Tracking Error. But for the funds QQQ and VGT, they follow their index's investment weights closely.

This tells us that the funds that replicate their index more closely have a lower tracking index like QQQ and VGT. Whereas ROBO, AIA, and AIRR have a higher Tracking Error due to their different approach in replicating the index.



Risk Levels and Performance of the Funds:

The risk adjusted return measure, the sharpe ratio, for the ETFs and the indices is calculated. VGT has the highest annual sharpe ratio among all five funds and MSCI (the tracked index by VGT) has the highest sharpe ratio among other tracked indices. The respective sharpe ratios of both are 0.998 and 0.927. However, VGT outperformed its index with a higher sharpe ratio making it an appealing investment regarding the studied ratio measure.

To estimate the performance and observe other risk measures for a fund, we have used the single factor model known as Capital Asset Pricing Model to estimate the Alpha and Beta of the fund. The values for Alpha and Beta are calculated by regressing the fund's excess returns over the market's excess returns. The excess return is the difference between the fund return and the quoted risk-free rate for that periodicity. Here we consider the 13-week Treasury bill rate as the risk-free rate. Since this rate is quoted on an annual basis, to convert it to a weekly rate we divide it by 52.

Once the excess returns are calculated, the least mean squared regression model is used to find the best fit line for the fund and market excess return. We considered S&P 500 and the Dow Jones Technology Index as the market representative.

Beta is the risk of the fund which is measured as the risk relative to the market risk. It is also called as systematic risk which cannot be diversified. It is the slope of the regression line used in the factor model. Thus, if a fund's beta is 1.0 it means that the risk of the fund is equal to

the market risk. As you can see the result from the regression model above, 4 out of 5 funds have their beta over 1.0, which indicates that their risk is more than the market risk of S&P 500 Index. It is used as the multiply factor in the CAPM to calculate the required rate of return. We can see that AIA has a beta which is less than 1.0, this indicates that the relative risk of AIA ETF is less than the S&P 500. One of the reasons for that can be, as we have seen that AIA invest 99% of its funds in Large-cap which are safer investments with respect to volatility.

The ETFs risk compared to the Dow Jones Technology Index is relative less as all five fund's betas is less than 1.0. Since DJUSTC consists of wide range of technology companies it takes more risk than the ETFs we have selected as they invest majorly in AI and automation companies.

Looking at “alpha” is a measure that is used to check if the fund is outperforming the market. In simple terms, if the alpha is positive then the fund beats the market and if it is negative it underperforms than the market. Since ETFs track an index, as fund managers, they use different strategies to get a return higher than the market for its investors. This higher rate of return is also called as the abnormal rate of return which is measured as alpha. It is the intercept of the regression line.

These are the following values for alpha for our funds vs the market:

	S & P 500 Index	Dow Jones Tech Index
ROBO	-0.0004042262	-0.0010292078
AIA	-0.0001075171	-0.0007692377
AIRR	-0.0010383895	-0.0013581653
QQQ	0.0011885163	0.0002884379
VGT	0.0014883260	0.0005149186

As you can observe, the ETFs ROBO, AIA and AIRR have a negative alpha indicating that they do not beat either of the market returns and are underperforming. AIRR ETF has the lowest alpha and as we have seen throughout the analysis, AIRR has been a volatile fund which the highest tracking error and a low degree of index replication. The ETFs QQQ and VGT have a positive alpha, thus indicating that it outperforms the market. The VGT ETF has the highest alpha and the analysis has backed that as it has the lowest expense ratio thus having a stable and low tracking error.

Moving on to covariance, the covariance is a measure of linear dependence between the two random variables. To obtain a relationship between ETFs and conventional funds,

covariance and correlation are very helpful tools. Covariance determines joint variability of two random variables. Where a greater value of a one variable correlate positively with a greater value of another variable. For instance, ROBO is correlated with one-self will give us a greatest value than ROBO with another ETFs. This behavior is being seen in all the other ETFs also, that they project higher covariance with their own variable. In contrast, AIRR has a higher covariance (0.0008086) amongst other ETFs, which might reflect that AIRR with itself corresponds more positively compare to others with themselves. QQQ, VGT and AIRR are the ETFs which positively and highly corresponds with S&P 500 in our portfolio following behind ROBO and AIA. Moreover, a covariance with DJUSTC also depicts a similar relation as with S&P 500 where VGT is on a top.

Correlation gives a better idea comparatively to a covariance. Where, if we investigate the measure very closely, we find that ROBO has good correlation of 0.8396 with S&P 500 even though its covariance is still less. With increases in correlation risk will also be rise, as AIA is a least correlated with indices S&P 500 and DJUSTC, when an index perform worst it will have less effect on AIA then VGT and QQQ because VGT is highest related to DJUSTC and QQQ performs better with S&P 500.

As every ETFs best correlate with themselves by having 1.0 correlation constant. Moreover, VGT and QQQ both has second highest correlation between each other as they both are very close to follow indices. Except for AIRR and AIA shows close relationship whereas ROBO has a higher correlation with VGT. Here, we can imply that correlation of one with another doesn't mean that other ETFs have also the same relationship but might be different.

Analysis of Variance (ANOVA):

ANOVA is a collection of variation statistical model which separate data variance into different components for a test, which is developed by a statistician and evolutionary biologist Ronald Fisher. Basically, ANOVA model is more conservative as it consists fewer type 1 error. According to our ANOVA test we have market excess and residuals of between ETFs and Indices. Residuals represent a left part after the F-test between two group which shows how

much ETF is influenced by Indices. Moreover, our P-values are minimal which determines the rejection of a null hypothesis, considering alpha value 0.05.

F-test evaluates explained variables over unexplained variables. Here, we explained variables shows variability between group and unexplained variables are within group variability. Higher F-test value determines effect of indices on ETFs is higher while having less variability shows that ETFs are more depended on Indices then its past record.

One-way ANOVA compares a separate means model to equal means model and Two-way model compares regression model to equal means model. Where the difference between F values of both will help us to determine lake of fit F-test.

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

Residual shows us the undefined variation in a response with all the contained features in our dataset (Here, we have two features). There are two assumptions about residuals. First one is the random variation around a line (fitted values) same for a every type of means of a factor or we can say across a covariate range. Second assumption is residuals having variance are independent of each other. Last one is generating a normal distribution by residuals.

According to our analysis we are focusing on higher F-test value because that will show us whether a presence of the influence by Index on ETF and lower value shows less impact on ETF by an Index. Here, we have QQQ and VGT has way high F-value while rejecting null hypothesis and having a less residual value, whereas AIA and AIRR somehow shows less dependency having high residual value. If we look into a graph, (in the appendix) it seems that residuals “bounce randomly” around 0 line which means that an assumption that the relationship is linear is reasonable. The residuals are not forming a complete horizontal line that suggest us that the variance of the error terms is equal. And there still some outliers standing from the basic pattern of residuals. To analyze more closely we are taking the square root of standardize residuals, which reflects that some residual is negative. Our model shows that it is correct by not having nonlinearity and heteroscedasticity.

$$Y_j - Y_{j(I)} / \text{MSE} \cdot p \text{ (Cook's distance vs leverage)}$$

Here, Y_j = is the fitted value for the j observation; $Y_{j(I)}$ = is the fitted value for the j observation without including the i-th observation in the data that will generate the model; p is the number of parameters in the model; MSE i.e. the mean squared error of the model. Cook's

distance measures the influence on a point or a region with respect to other related data of (x_i, y_i) is removed. Leverage measures how much one feature is leveraged by its surrounding features in our data. According to our graph, we can analyze that most of a point has less leverage and less Cook's distance, which means they are not much influenced whereas the other who has a higher Cook's distance and less leveraged.

Our graphs (attached in the appendix) shows most a point fall from a range of 0.01 to 0.06 of Cook's distance where most of them having higher 242 and 257 Obs. Number. The less Cook's number means after a removing impact of indices values how far a prediction point is far forming a worst condition of ETFs. Higher Hat values increase the Cook's distance. Suppose if a Cook's distance value is 1 then those are highly influenced points on a graph and to indicate high leverage points we calculate leverage values greater than 2 times the number of predictors devices by the number of features. High leverage indicates a point whose value is far from its average which can greatly influence the model. According to our model, we have less points indicating high leverage. As we have less data-point showing high leverage out ANOVA model, it finally can be concluded that we have a good compatible or fit model for a further implementation.

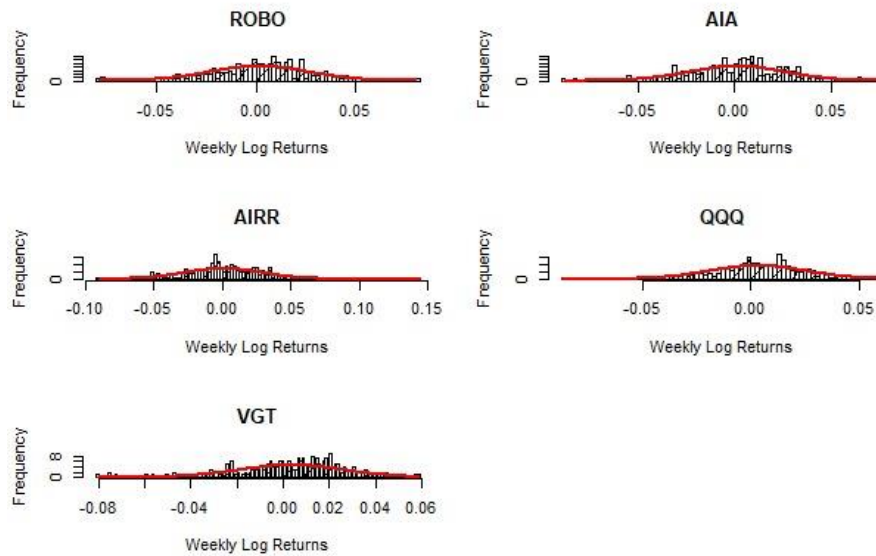
Conclusion:

To sum up our analysis, with respect to an investor's approach, VGT and QQQ have shown merit to invest in. They track their indices very closely and also outperform it. Comparing QQQ with VGT we have found that QQQ stays aligned to an index thus it is unlikely of getting abnormal returns. Our analysis meets a requirement for a Cook's distance and a leverage graph with falling more points to less leverage field.

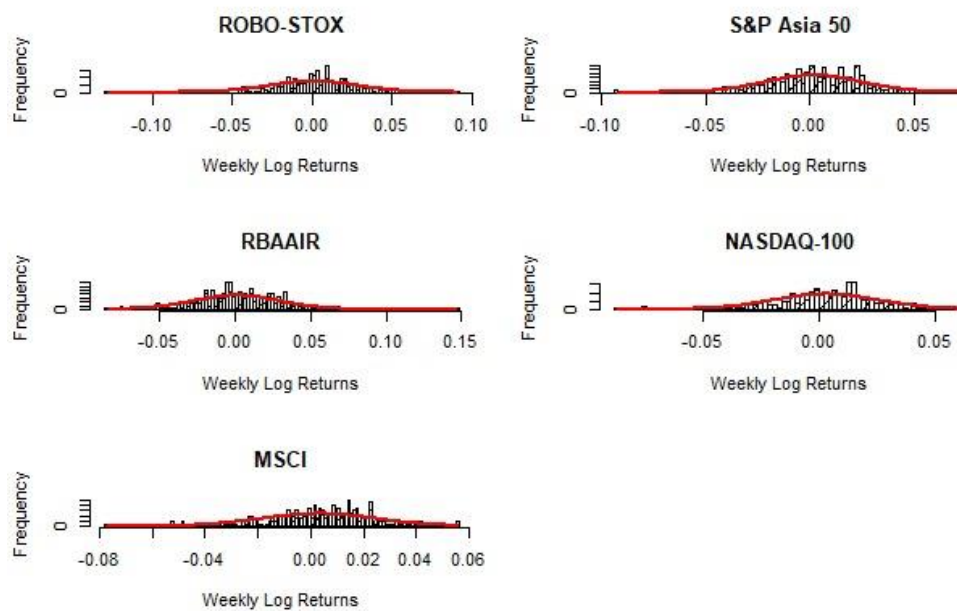
Appendix

Output for R Code Analysis:

ETF Log Returns Distribution



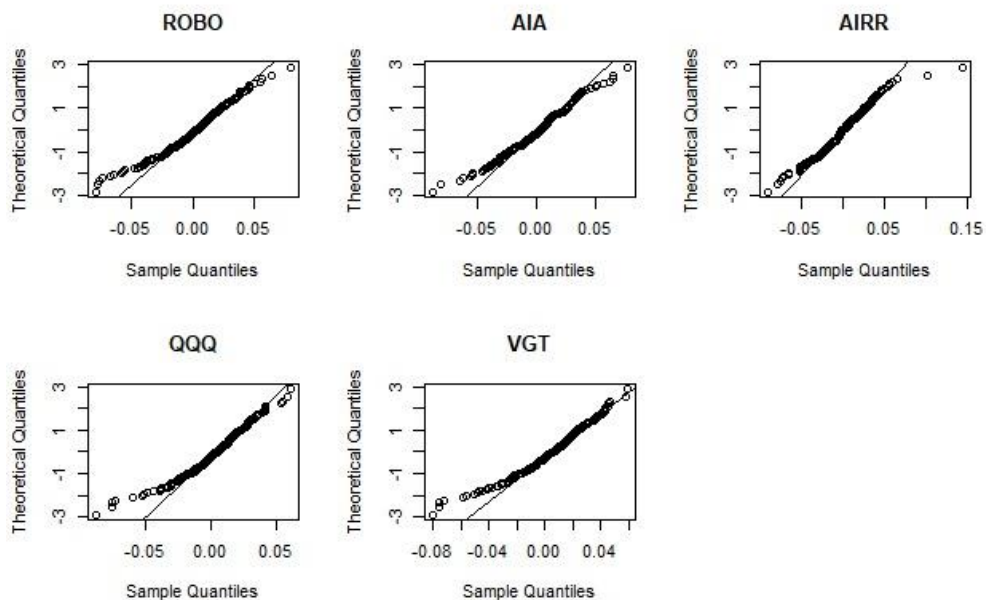
ETF Index Log Returns Distribution



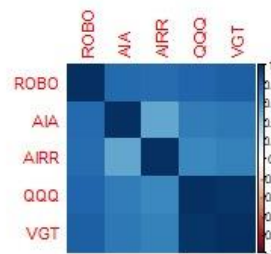
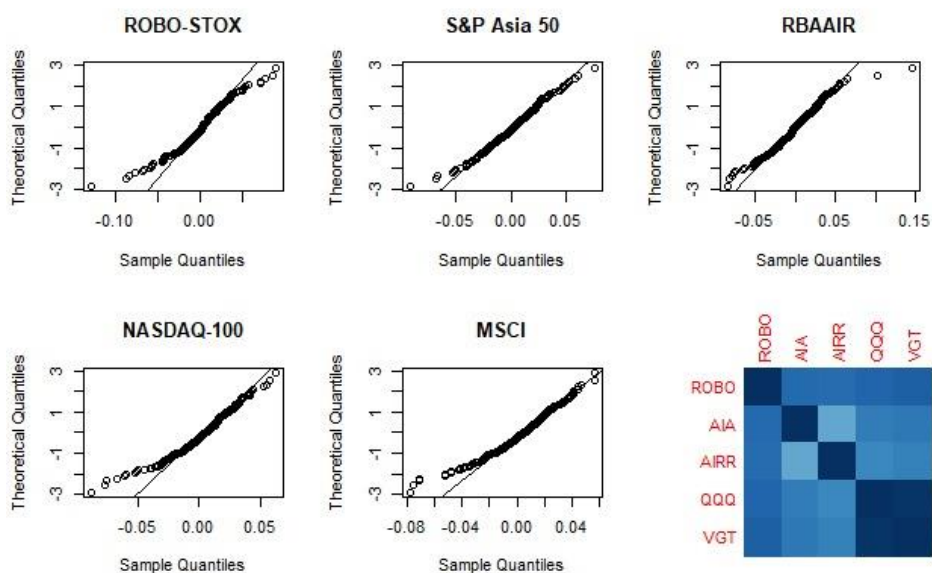
Market Log Returns Distribution



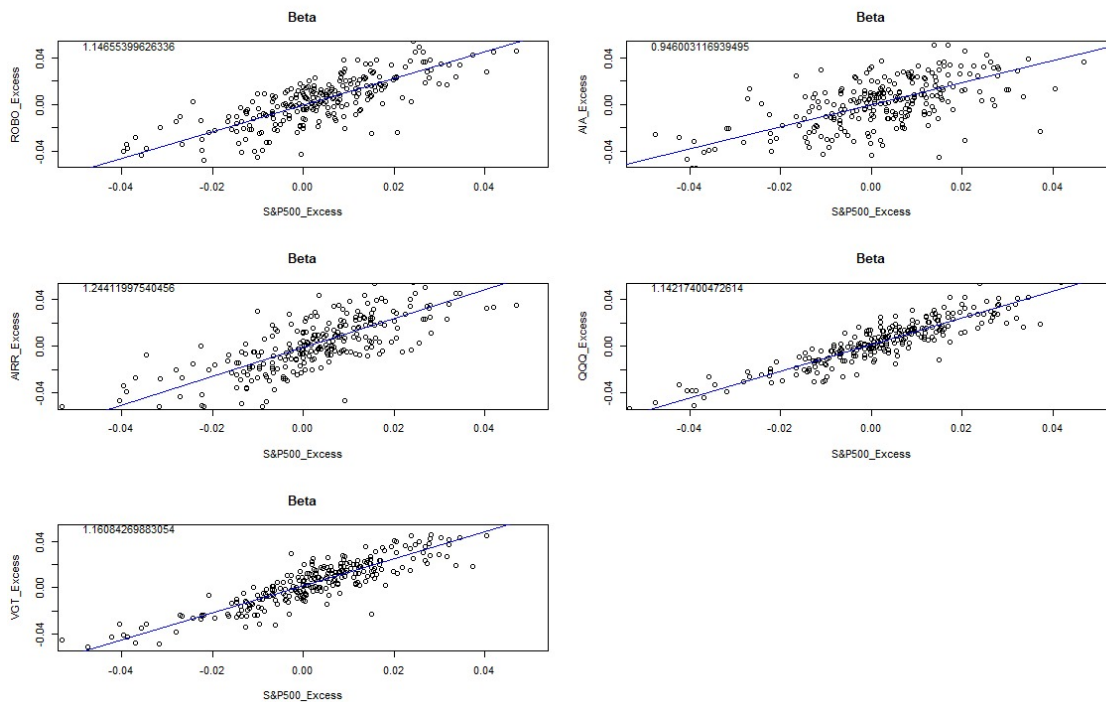
Shapiro Test on ETFs



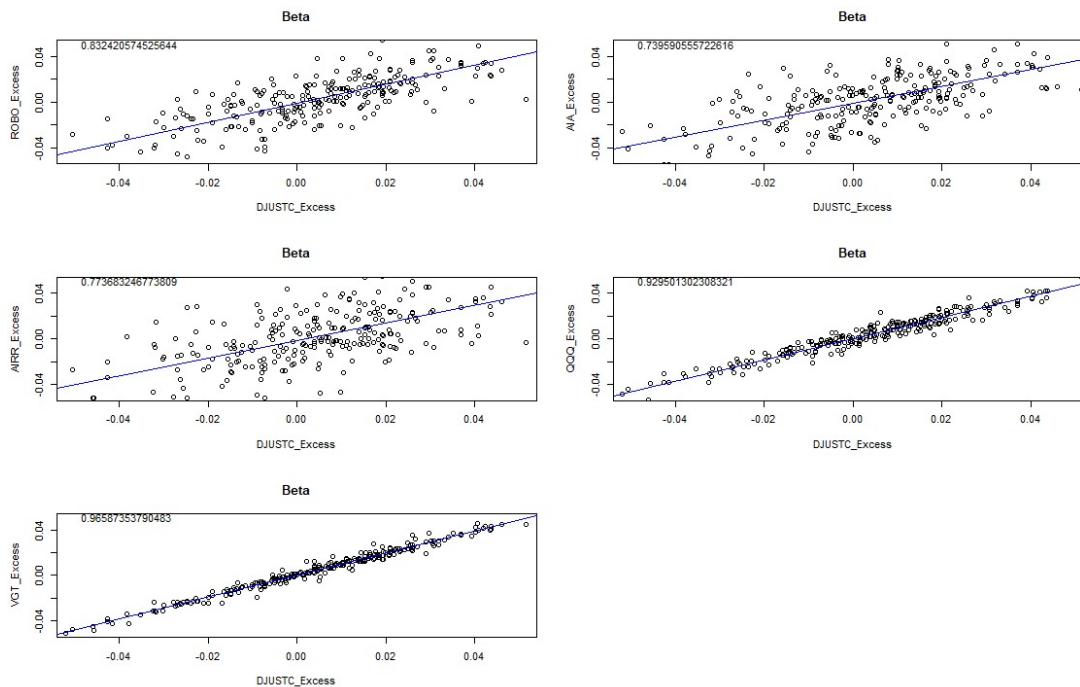
Shapiro Test on ETF Indices and Correlation Matrix of ETFs



CAPM Regression of ETF Excess Returns on S&P 500 Excess Returns



CAPM Regression of ETF Excess Returns on Dow Jones Technology Index



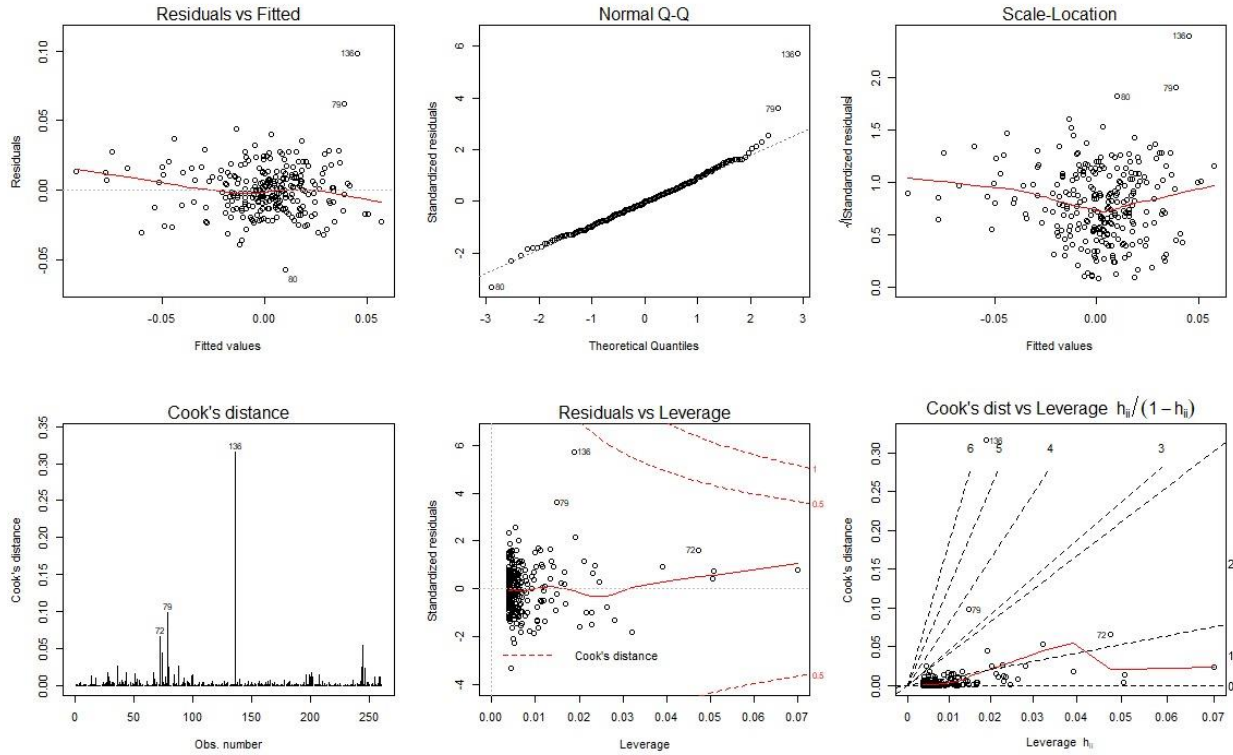
The figure displays six diagnostic plots for a linear regression model, arranged in a 2x3 grid. The top row contains three plots: 'Residuals vs Fitted', 'Normal Q-Q', and 'Scale-Location'. The bottom row contains three plots: 'Cook's distance', 'Residuals vs Leverage', and 'Cook's dist vs Leverage $h_i/(1-h_i)$ '.

- Residuals vs Fitted:** A scatter plot showing residuals on the y-axis (ranging from -0.04 to 0.06) against fitted values on the x-axis (ranging from -0.08 to 0.04). A red smoothing line is shown. Points 2390, 212, and 240 are labeled.
- Normal Q-Q:** A plot of standardized residuals on the y-axis (ranging from -2 to 4) against theoretical quantiles on the x-axis (ranging from -3 to 3). A dashed diagonal line represents the expected normal distribution. Points 2390 and 240 are labeled.
- Scale-Location:** A plot of the square root of the absolute value of standardized residuals on the y-axis (ranging from 0.0 to 2.0) against fitted values on the x-axis (ranging from -0.08 to 0.04). A red smoothing line is shown. Points 2390, 2400, and 212 are labeled.
- Cook's distance:** A bar plot showing Cook's distance on the y-axis (ranging from 0.00 to 0.08) against observation number on the x-axis (ranging from 0 to 250). Points 98, 239, and 240 are labeled.
- Residuals vs Leverage:** A scatter plot showing standardized residuals on the y-axis (ranging from -4 to 4) against leverage on the x-axis (ranging from 0.00 to 0.07). A red smoothing line and dashed Cook's distance contours are shown. Points 9239, 99, and 240 are labeled.
- Cook's dist vs Leverage $h_i/(1-h_i)$:** A scatter plot showing Cook's distance on the y-axis (ranging from 0.00 to 0.08) against leverage h_i on the x-axis (ranging from 0 to 0.07). Dashed lines represent Cook's distance contours, and a red smoothing line is shown. Points 4, 9239, 3, 2, and 99 are labeled.

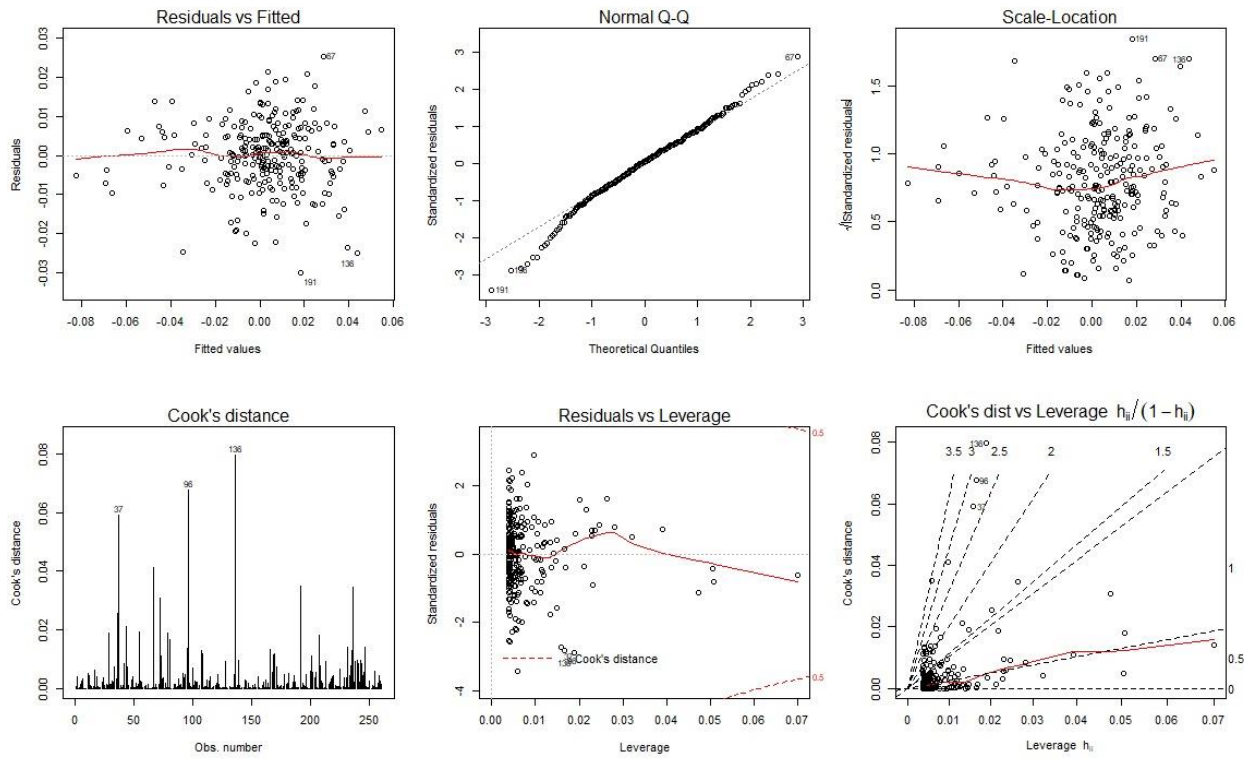
The figure displays six diagnostic plots for a linear regression model, arranged in a 2x3 grid. The plots are used to assess the model's assumptions and identify influential observations.

- Top Left: Residuals vs Fitted**
 - Y-axis: Residuals (ranging from -0.06 to 0.04)
 - X-axis: Fitted values (ranging from -0.06 to 0.04)
 - Shows a scatter of points around a horizontal red line at zero. Points 136, 191, and 260 are labeled.
- Top Middle: Normal Q-Q**
 - Y-axis: Standardized residuals (ranging from -3 to 3)
 - X-axis: Theoretical Quantiles (ranging from -3 to 3)
 - Points follow a diagonal line, indicating approximate normality. Points 136, 191, 240, and 260 are labeled.
- Top Right: Scale-Location**
 - Y-axis: $\sqrt{|Standardized residuals|}$ (ranging from 0.0 to 1.5)
 - X-axis: Fitted values (ranging from -0.06 to 0.04)
 - Shows a scatter of points around a red curve. Points 136, 191, 240, and 260 are labeled.
- Bottom Left: Cook's distance**
 - Y-axis: Cook's distance (ranging from 0.00 to 0.15)
 - X-axis: Obs. number (ranging from 0 to 250)
 - Shows a bar plot of Cook's distance for each observation. Points 72, 136, and 260 are labeled.
- Bottom Middle: Residuals vs Leverage**
 - Y-axis: Standardized residuals (ranging from -4 to 3)
 - X-axis: Leverage (ranging from 0.00 to 0.07)
 - Shows a scatter of points with a red curve. Points 136, 240, and 260 are labeled.
- Bottom Right: Cook's dist vs Leverage $h_i/(1-h_i)$**
 - Y-axis: Cook's distance (ranging from 0.00 to 0.15)
 - X-axis: Leverage h_i (ranging from 0 to 0.07)
 - Shows a scatter of points with a red curve and dashed lines representing Cook's distance contours (0.5, 1, 1.5, 2, 2.5, 3, 3.5). Points 72, 136, 240, and 260 are labeled.

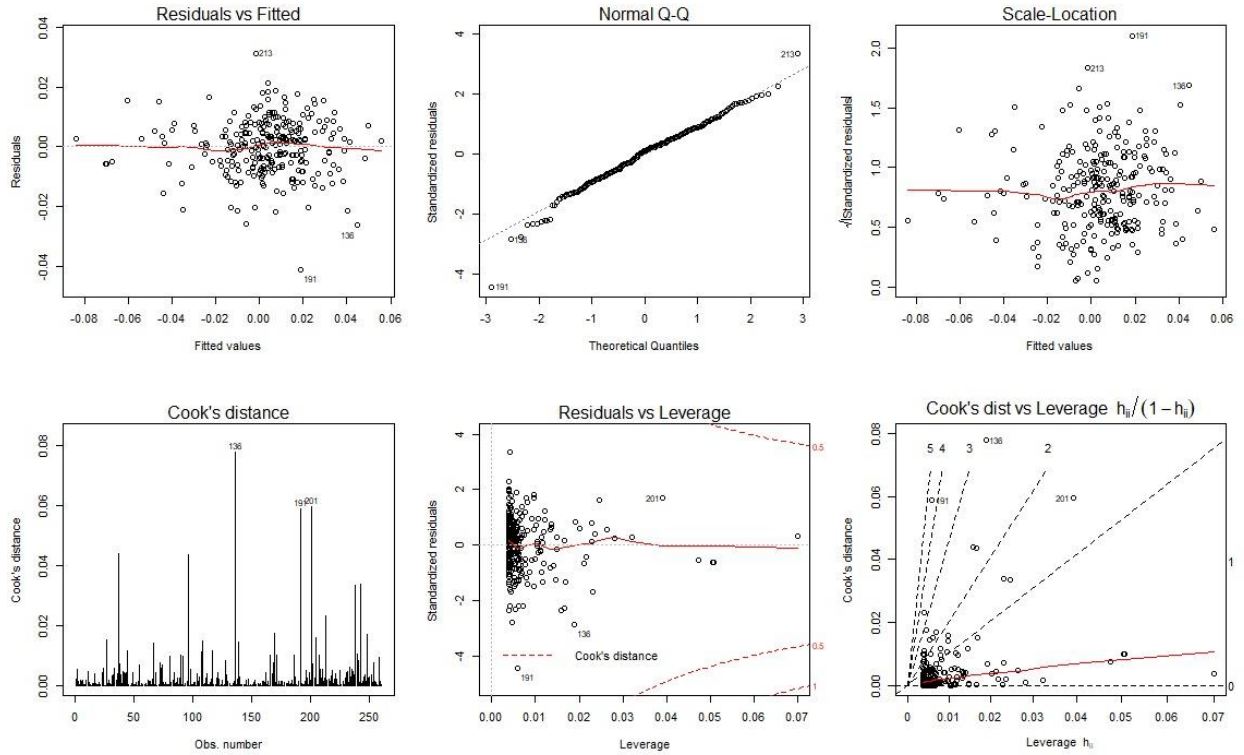
ANOVA: AIRR vs S&P



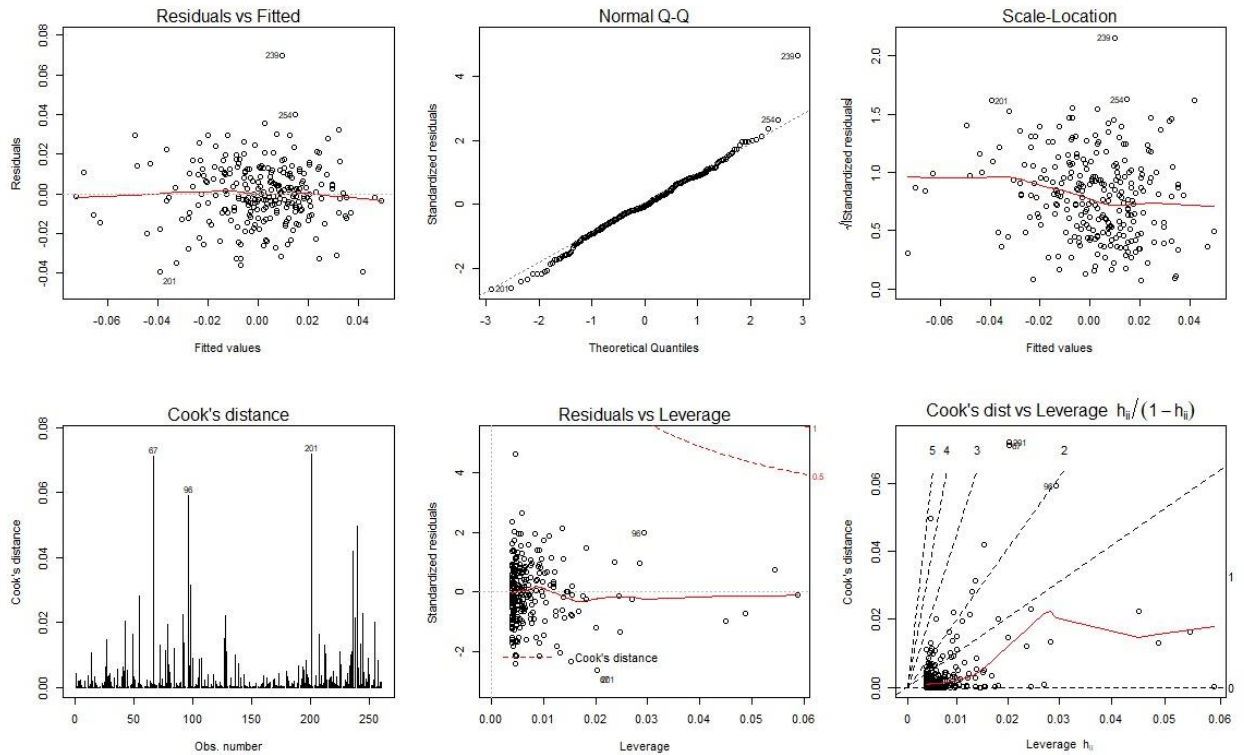
ANOVA: QQQ vs S&P



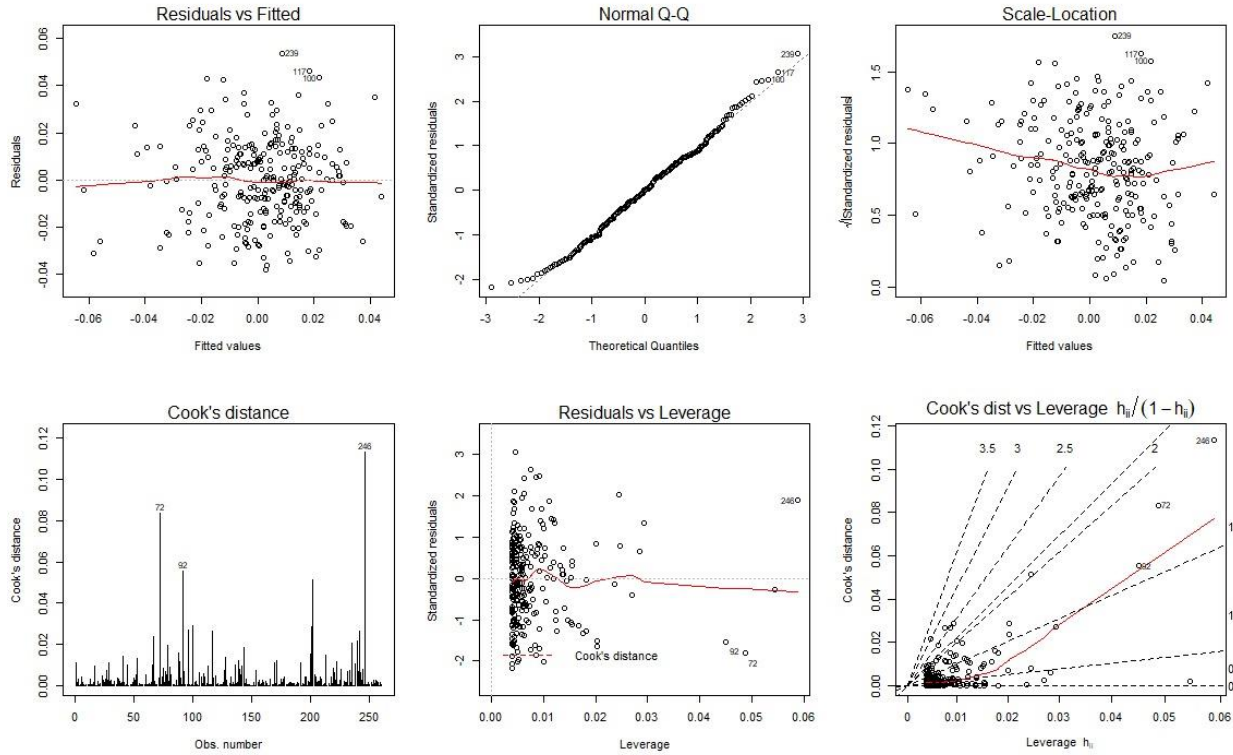
ANOVA: VGT vs S&P



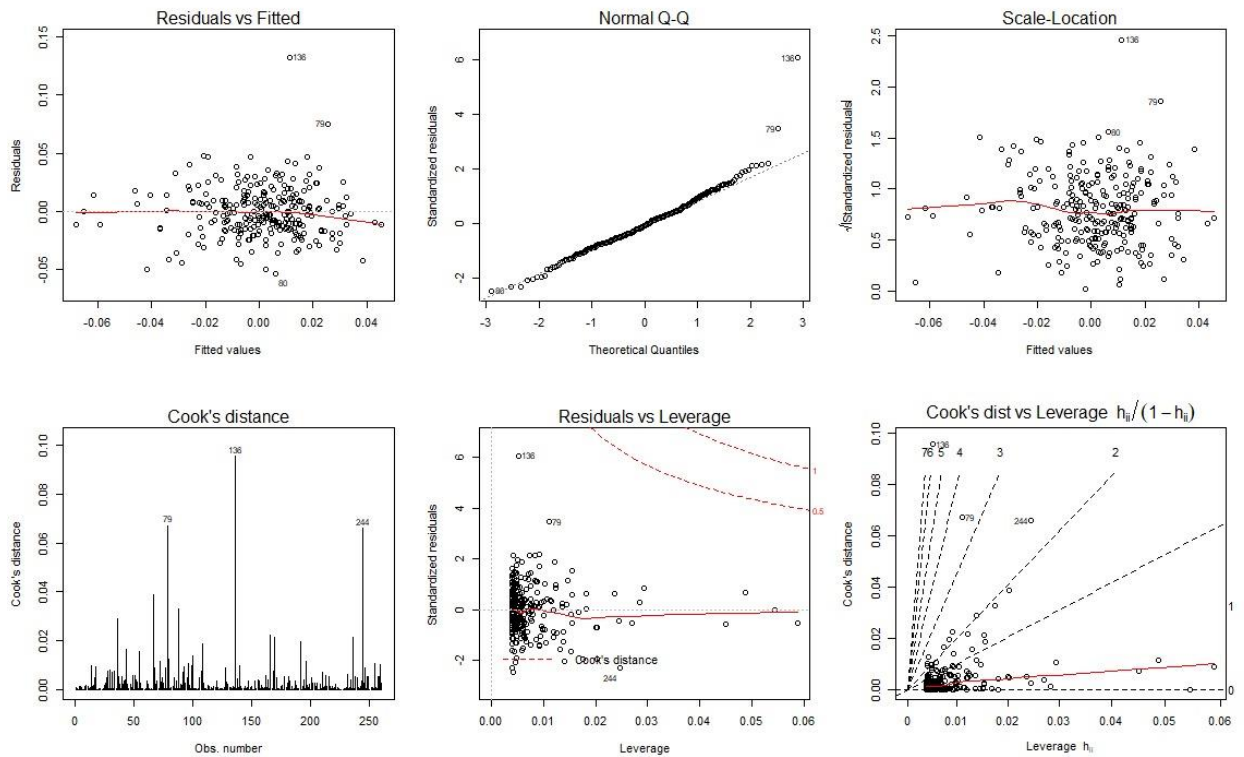
ANOVA: ROBO vs DJUSTC



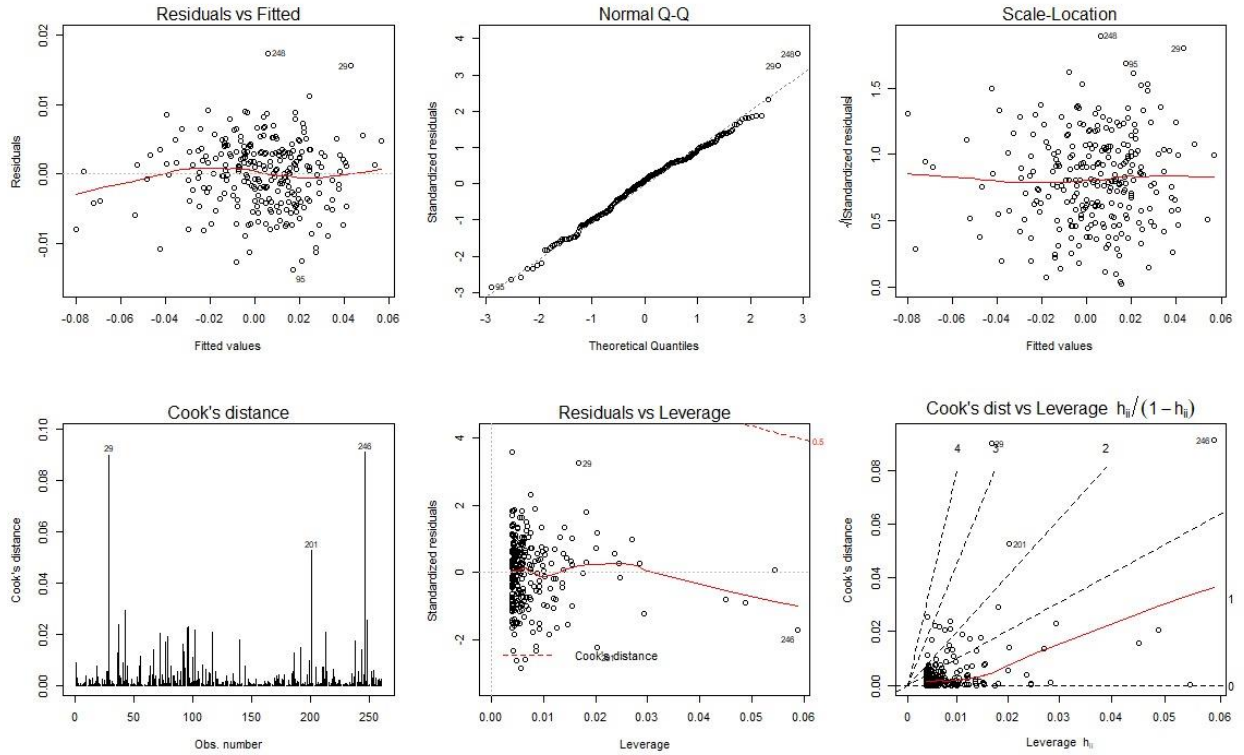
ANOVA: AIA vs DJUSTC



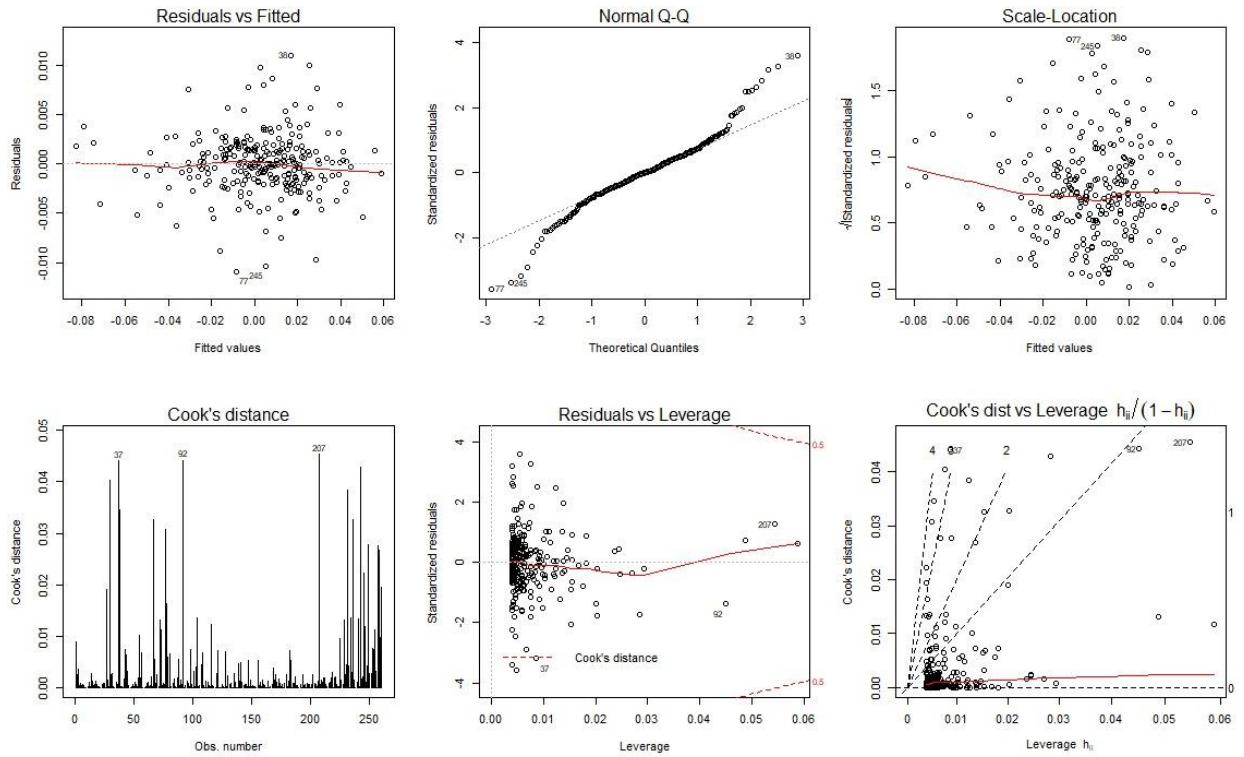
ANOVA: AIRR vs DJUSTC



ANOVA: QQQ vs DJUSTC



ANOVA: VGT vs DJUSTC



References

n.d. *Analysis of Variance*. https://en.wikipedia.org/wiki/Analysis_of_variance

n.d. *Anova - Investopedia*. <https://www.investopedia.com/terms/a/anova.asp>

Chen, James. 2019. *Tracking Error*. February 21.
<https://www.investopedia.com/terms/t/trackingerror.asp>

ETF.com. n.d. *Understanding Tracking Difference and Tracking Error*. <https://www.etf.com/etf-education-center/21030-understanding-tracking-difference-and-tracking-error.html>.
n.d. *fidility.com*.

Gafar Matanmi Oyeyemi, Abdulwasiiu Bukoye, Imam Akeyede. 2015. "Comparison of Outlier Detection Procedures in Multiple Linear Regressions." *American Journal of Mathematics and Statistics*. <http://article.sapub.org/10.5923.j.ajms.20150501.06.html>.