# AI ETF Fund Analysis and Portfolio Management

Sasi Nakchatree

Norberto J. Rancharan

Rajat D. Sudagade

The University of Texas at Dallas

Author Note

Department: Financial Engineering and Risk Management (JSOM)

Guided By: Dr. Liping Ma

Contact: sasi.nakchatree@utdallas.edu, norberto.rancharan@utdallas.edu

rajat.s@utdallas.edu

Abstract

An Exchange-Traded Fund (ETF) is an investment type fund that is, much like stocks, traded on stock exchanges.[1][2] It can hold various types of assets like stocks, commodities, or bonds and it generally operates with a mechanism that keeps its trading values close to its net asset value.[3]

ETFs are considered to be an attractive sort of investment because they have low costs, they have lower tax rates, and they also have stock-like features. [4][5] ETFs also act like a mutual fund in the way that it can be bought or sold at the end of each trading day for its net asset value (NAV), but it also has features of a closed-end fund and hence can be traded throughout the day at prices that can be greater or lesser than its NAV.[3]

This project aims towards analyzing the performance of 5 large AI ETFs and assess the various types of risks that an individual can come across while dealing with them.

*Keywords:* ETF, Mutual Fund, Stocks, AI

AI ETF Fund Analysis and Portfolio Management

## I. Introduction

An Exchange-Traded Fund (ETF) is an investment type fund that is, much like stocks, traded on stock exchanges.[1][2] It can hold various types of assets like stocks, commodities, or bonds and it generally operates with a mechanism that keeps its trading values close to its net asset value.[3]

Exchange Traded Fund can be bought or sold directly, or to an authorized broker-dealer. Distributors of ETFs enter into agreements with these dealers and create a unit consisting of thousands of ETF shares which are then exchanged in-kind with underlying securities. Participants can then invest in these shares on a long term basis or act as market makers to exchange the units in orders to provide liquidity of the shares. This complicated procedure helps keep the intraday market price of the ETF very close to the net asset value of the underlying assets.

ETFs are considered to be an attractive sort of investment because they have low costs, they have lower tax rates, and they also have stock-like features. [4][5] ETFs also act like a mutual fund in the way that it can be bought or sold at the end of each trading day for its net asset value (NAV), but it also has features of a closed-end fund and hence can be traded throughout the day at prices that can be greater or lesser than its NAV.[3]

ETFs were introduced in the United States in 1993 (and 1999 in Europe), and until 2015 they have acquired more than US$2 trillion worth of investments. Currently ETF offer more than 1,800 different products that cover almost every type of market and trading strategies that can be thought of.

This project aims towards analyzing the performance of large AI ETFs and assess the various types of risks that an individual can come across while dealing with them. The entire project is based on the R programming language and tracks the performance of the following ETFs:

**Invesco QQQ Trust (QQQ)**

Formerly known as Powershares QQQ, this ETF tracks a modified-market-cap-weighted index of 100 NASDAQ-listed stocks. QQQ is considered to be one of the best and most traded ETFs in the world. According to its index, QQQ only invests in nonfinancial stocks that are listed on NASDAQ, and ignores everything else. While QQQ is highly exposed to tech companies, it is not considered as a tech fund. Instead, QQQ provides investors with a tasteful mix of tech, growth, and large-cap exposure.

**Technology Select Sector SPDR Fund (XLK)**

XLK exposes us to the tech sector in United States, but it only tracks companies that are listed in the S&P 500 portfolio. It has a lot of big, tech-related names covered along with a few newcomers that are working in financial payment processing and telecom. XLK avoid small and mid caps, hence presenting us with a lower volatility rate. XLK is one of the cheapest but largest funds in the tech sector and is heavily attractive to liquid-active traders.

**Invesco NASDAQ Internet ETF (PNQI)**

This ETF invests in a huge number of US-listed internet companies. It also includes international companies that are listed on US exchanges, such as China-based Baidu.

**First Trust Dow Jones Internet Index Fund (FDN)**

FDN tracks 40 of the biggest US-based internet companies such as Amazon, Facebook, and Google. Among all the internet based ETFs, FDN does a great job at trading while staying within a reasonable cost range

**iShares U.S. Technology ETF**

This Exchange-Traded Fund fund tracks a market-cap weighted index of U.S. technology companies and offers a broad and representative exposure to these companies. The ETF's assets are concentrated heavily among its top securities. As a well-run and extremely efficient fund, IYW comes with a reasonable expense ratio and very close tracking.

## II. Initial Setup

We begin by fetching our primary dataset - the 5 previously mentioned datasets. This can be done by using the **getSymbols** function from the **quantmod** R package to create a time series for each of these datasets. We are fetching all the **weekly** data of the ETF from **January 1, 2015** to **January 1, 2019** as shown below:

```
symbols=c("QQQ", "XLK" , "PNQI" , "FDN" , "IYW")
getSymbols(symbols,from="2015-01-01",to="2019-01-01",periodicity="weekly",
           return.class="ts")

mydat<-cbind(QQQ[,6], XLK[,6], PNQI[,6], FDN[,6], IYW[,6])
colnames(mydat)<-c("QQQ", "XLK" , "PNQI" , "FDN" , "IYW")
```

This will give us a new time series called **mydat** that contains the adjusted prices of these ETFs. Given below are some of the initial values present in this new time series.

| QQQ | XLK | PQNI | FDN | IYW |
|---|---|---|---|---|
| 97.34110 | 37.77940 | 64.97489 | 59.41 | 97.58478 |
| 96.95695 | 37.79805 | 64.32500 | 59.00 | 97.73807 |
| 98.09017 | 38.05931 | 65.57471 | 59.54 | 98.15003 |
| 96.91856 | 37.45085 | 65.54472 | 59.53 | 96.72258 |
| 98.79123 | 38.11528 | 69.27432 | 61.32 | 98.25540 |
| 100.62551 | 38.01102 | 67.71422 | 62.52 | 100.60259 |

An ETF is designed to track a specific index. It can be a bond index, a stock index, or a price index. Hence we need to fetch the data of an index over the same time period. For this project, we are fetching the data of two general market indexes - The S&P 500 Index (GSPC) and the Dow Jones US Technology Index (DJUSTC). This can be done using the **getSymbols** function as shown below:

```
symbols=c("^GSPC", "^IRX")
getSymbols(symbols,from="2015-01-01",to="2019-01-01",periodicity="weekly",
           return.class="xts")

DJI = read.csv(file="DJUSTC.csv", header=TRUE,  sep=",")

markdat = cbind(GSPC[,6],DJI[,1])
colnames(markdat)<-c( "GSPC", "DJI")

DJINET = read.csv(file="DJINET.csv", header=TRUE,  sep=",")
IXTNTR = read.csv(file="IXTNTR.csv", header=TRUE,  sep=",")
NDX = read.csv(file="NDX.csv", header=TRUE,  sep=",")
QNET = read.csv(file="QNET.csv", header=TRUE,  sep=",")

ETFindex = cbind(DJINET[,1],IXTNTR[,1],NDX[,1],QNET[,1], DJI[,1])
colnames(ETFindex)<-c("DJINET", "IXTNTR", "NDX", "QNET","DJIX")
```

Along with the general market indexes GSPC and DJUSTC, we are also fetching the data of indexes that are being tracked by our 5 exchange traded funds. We are then creating a new time series object containing the **Adjusted Closing Prices** of these indexes. Given below are some of the initial values present in this new time series.

| DJINET | IXTNTR | NDX | QNET | DJIX |
|---|---|---|---|---|
| 267.29 | 1685.58 | 4213.28 | 383.4 | 23 |
| 259.97 | 1651.33 | 4142.14 | 398.3 | 29 |
| 273.45 | 1695.24 | 4278.14 | 398.0 | 29 |
| 297.38 | 1629.67 | 4148.43 | 389.0 | 8 |
| 275.94 | 1679.77 | 4228.68 | 394.5 | 19 |
| 287.33 | 1744.25 | 4384.03 | 409.5 | 53 |

## III. Analysis of Weekly Returns of ETFs

We have created a time series object containing the Adjusted Closing Price of each of the ETFs. Let's analyse the weekly returns of these ETFs using the general formula of:

$$r_t = log(P_t) - log(P_{t-1})$$

From this formula, the return at time *t* is calculated by taking the difference between the log values of the adjusted closing prices at time *t* and *t-1*. In R, we will do this as shown below:

```
wkQQQret = diff(log(mydat[,1]))
wkXLKret = diff(log(mydat[,2]))
wkPNQIret = diff(log(mydat[,3]))
wkFDNret = diff(log(mydat[,4]))
wkIYWret = diff(log(mydat[,5]))

ret = cbind(wkQQQret,wkXLKret,wkPNQIret,wkFDNret,wkIYWret)
colnames(ret) = c("QQQ", "XLK" , "PNQI" , "FDN" , "IYW" )
```

We are calculating the weekly returns for each of ETFs and saving the values inside a new time series object called **ret**. Here are some of its initial values:

| QQQ | XLK | PNQI | FDN | IYW |
|---|---|---|---|---|
| -0.00395421 | 0.000493534 | -0.00727844 | -0.00692511 | 0.001569678 |
| 0.011620061 | 0.006888351 | 0.019241746 | 0.009110928 | 0.00420605 |
| -0.01201618 | -0.01905772 | -0.00045746 | -0.00016800 | -0.01465035 |
| 0.019137829 | 0.020527321 | 0.026045918 | 0.029625684 | 0.01572351 |
| 0.018396988 | 0.023228839 | 0.006517658 | 0.019380452 | 0.023607722 |
| 0.022460083 | 0.021297506 | 0.028384024 | 0.033656329 | 0.02714933 |

Similarly, we have calculated the weekly returns of the two general market indexes and the indexes that are tracked by the ETFs.

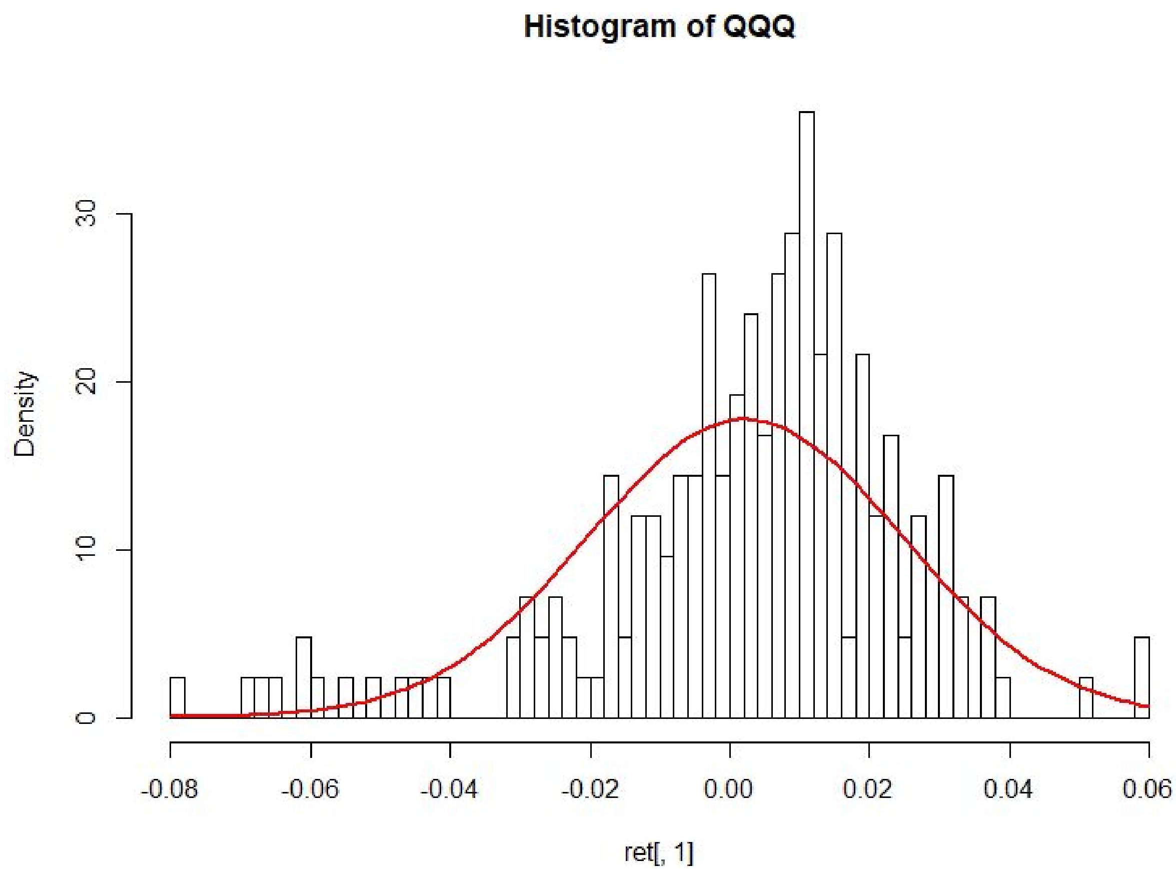
```
wkGSPCret = diff(log(markdat[,1]))
wkDJIret = diff(log(markdat[,2]))
bench.ret = na.omit(wkGSPCret)

markret = cbind(wkGSPCret,wkDJIret)
colnames(markret) = c("GSPC","DJI")
markret=na.omit(markret)

wkDJINETret = diff(log(ETFindex[,1]))
wkIXTNTRret = diff(log(ETFindex[,2]))
wkNDXret = diff(log(ETFindex[,3]))
wkQNETret = diff(log(ETFindex[,4]))
wkDJIXret = diff(log(ETFindex[,5]))


indexret = cbind(wkDJINETret,wkIXTNTRret,wkNDXret,wkQNETret,wkDJIXret)
colnames(indexret) = c("DJINET", "IXTNTR", "NDX", "QNET","DJIX")
```

Finally, when we plot the weekly returns of ETFs, we can see that the ETFs follow an almost normal distribution pattern. Here is the histogram plot of the Invesco QQQ Trust ETF:




IV. Performance of Funds

The project now takes a look at the correlation, variance and covariance of the ETFs' weekly returns. The R programming language has provided us with a few simple function that can calculate the variance, covariance and the correlation for us.

```
var(ret) //variance and covariance
cor(ret) // correlation
```

The function **var()** calculates the variance and the covariance of the weekly returns of the ETFs. The function outputs a matrix of numerical values, where the elements in the principal diagonal are the variances of the ETFs, and the rest of the elements represent the covariances of the ETFs with each other.

| | QQQ | XLK | PNQI | FDN | IYW |
|---|---|---|---|---|---|
| QQQ | 0.0005031 | 0.0004833 | 0.0005751 | 0.0005672 | 0.0005039 |
| XLK | 0.0004833 | 0.000486 | 0.0005352 | 0.0005353 | 0.000499 |
| PNQI | 0.0005751 | 0.0005352 | 0.0008026 | 0.0007566 | 0.000587 |
| FDN | 0.0005672 | 0.0005353 | 0.0007566 | 0.0007764 | 0.0005872 |
| IYW | 0.0005039 | 0.000499 | 0.000587 | 0.0005872 | 0.0005436 |

Similarly, the **cor()** function gives us the covariances between the exchange-traded funds. The principal diagonal elements of the output matrix will be equal to 1 because they represent the covariance of an ETF with itself.

| | QQQ | XLK | PNQI | FDN | IYW |
|---|---|---|---|---|---|
| QQQ | 1 | 0.9775 | 0.905 | 0.9076 | 0.9634 |
| XLK | 0.9775 | 1 | 0.8569 | 0.8715 | 0.9708 |
| PNQI | 0.905 | 0.8569 | 1 | 0.9584 | 0.8887 |
| FDN | 0.9076 | 0.8715 | 0.9584 | 1 | 0.9038 |
| IYW | 0.8715 | 0.9584 | 0.8887 | 0.9038 | 1 |

Next, we have calculated the risks of each of the exchange traded funds. Relative risk is the ratio of the probability of an outcome in a known data to the probability of the same outcome in some unknown data. In R, we measure the relative risks of the ETFs by calculating their mean, standard deviation, coefficient of variation, value at risk, kurtosis, and skewness as shown below:

```
RR = matrix(0,nrow = 7,ncol = 5,dimnames = list(c("Mean","SD","CV", "VaR", "Kurtosis", "Skewness"),
c("QQQ", "XLK" , "PNQI" , "FDN" , "IYW" )))
for(i in 1:5)
{
  RR[1,i] = mean(mydat[,i])
  RR[2,i] = sd(mydat[,i])
  RR[3,i] = sd(mydat[,i])/mean(mydat[,i])
  RR[4,i] = quantile( ret[,i], 0.05)
  RR[5,i] = kurtosis(mydat[,i])
  RR[6,i] = skewness(mydat[,i])
}
```

The output of above code snippet is a another matrix **RR** which contains a mathematical summary of the ETFs as shown below:

| | QQQ | XLK | PNQI | FDN | IYW |
|---|---|---|---|---|---|
| Mean | 130.1871 | 51.5797 | 96.6018 | 92.2072 | 132.1319 |
| SD | 27.1755 | 11.8136 | 23.5463 | 25.2852 | 31.275 |
| CV | 0.2087 | 0.229 | 0.2437 | 0.2742 | 0.2367 |
| VaR | -0.0445 | -0.0441 | -0.0494 | -0.049 | -0.0477 |
| Kurtosis | -1.195 | -1.2495 | -1.117 | -0.8835 | -1.2521 |
| Skewness | 0.4957 | 0.4554 | 0.487 | 0.642 | 0.4671 |

We can see here that the iShares ETF (IYW) has the highest mean and standard deviation. A high standard deviation indicates higher volatility of the ETF.

Finally, we compare the performance of the ETFs with the two general market indexes by calculating their mean as shown below:

```
ret_compar = cbind(wkQQQret, wkPNQIret, wkIYWret, wkXLKret, wkFDNret, wkGSPCret, wkDJIret)
ret_compar = na.omit(ret_compar)

pof = matrix(0,nrow = 1,ncol = 7,dimnames = list(c("mean"), c("QQQret", "XLKret" , "PNQIret" , "FDNret"
, "IYWret", "GSPCret", "DJIret" )))
for(i in 1:7)
{
  pof[1,i] = mean(ret_compar[,i])
}
```

The matrix obtained from the above code is shown below:

| | QQQret | XLKret | PNQIret | FDNret | IYWret | GSPCret | DJIret |
|---|---|---|---|---|---|---|---|
| mean | 0.0022 | 0.0025 | 0.00236485 | 0.00236453 | 0.0032 | 0.001 | 0.0088 |

## V. Tracking Error and Hypothesis Tests

Tracking Error is the difference between the return on the fund and the benchmark it is expected to track.

There are two ways to calculate the tracking error of a fund. The first method is where we can subtract the fund's returns from the return of the index that it is tracking.

The second method is one where we need to calculate the standard deviation of the difference between the returns of the fund and the index. We have used this method to calculate the tracking error of our ETFs.

```
QQQ_TE = sd(ret[,1]-indexret[,3])
XLK_TE = sd(ret[,2]-indexret[,2])
PNQI_TE = sd(ret[,3]-indexret[,4])
FDN_TE = sd(ret[,4]-indexret[,1])
IYW_TE = sd(ret[,5]-indexret[,5])
```

The tracking error of the ETFs is then obtained as shown below:

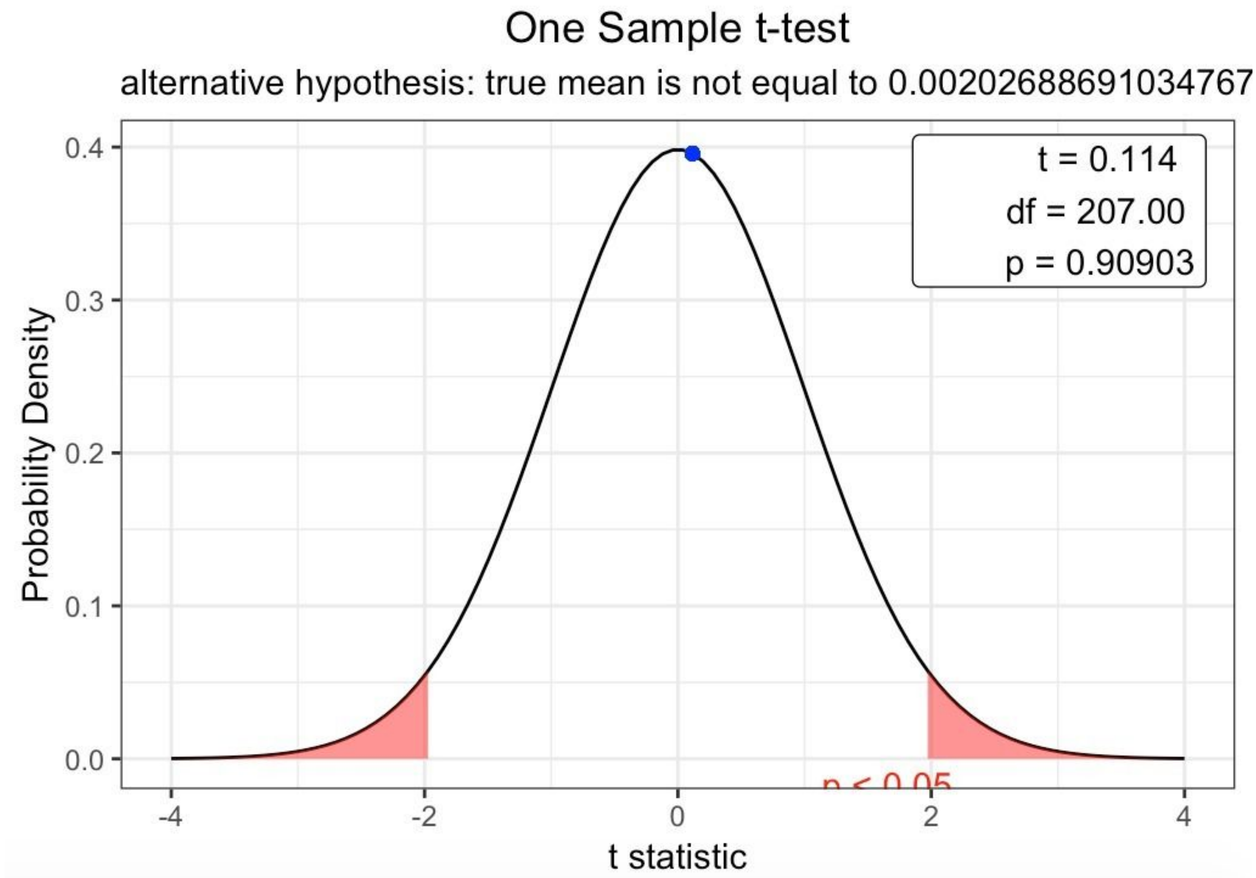
| QQQ | XLK | PNQI | FDN | IYW |
|---|---|---|---|---|
| 0.0232 | 0.0229 | 0.0265 | 0.026 | 0.024 |

Since the tracking error for each ETF is as low as 0.02, we can say that the ETFs are successful in keeping track of their respective indexes.

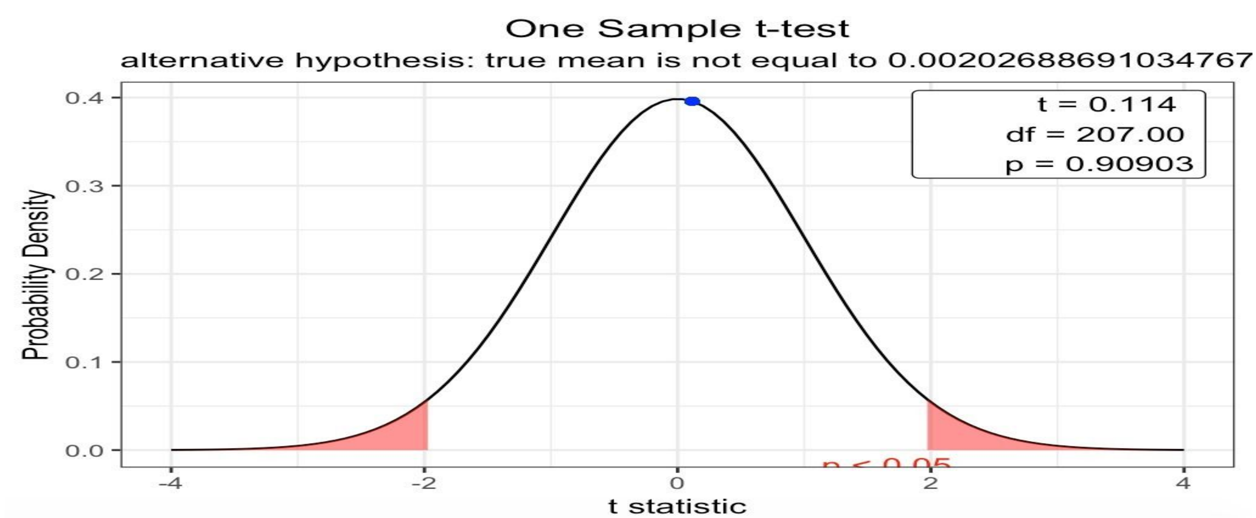
Also known as significance testing, hypothesis testing is a mathematical model that tests the claim, idea or hypothesis about a parameter of interest in a given population set using the data measured in a sample set. There are many ways to perform hypothesis tests. We have used T-Test method to perform hypothesis tests on our ETFs. T-Tests are used to determine if there is a significant difference between the means of two groups. We are using T-Tests to compare the mean of an ETF with the mean of an index as shown below:

```
QQQttest = t.test(ret[,1], mu = mean(indexret[,3]))
plot(QQQttest)
XLKttest = t.test(ret[,2], mu = mean(indexret[,2]))
plot(XLKttest)
PNQIttest = t.test(ret[,3], mu = mean(indexret[,4]))
plot(PNQIttest)
FDNttest = t.test(ret[,4], mu = mean(indexret[,1]))
plot(FDNttest)
IYWttest = t.test(ret[,5], mu = mean(indexret[,5]))
plot(IYWttest)
```

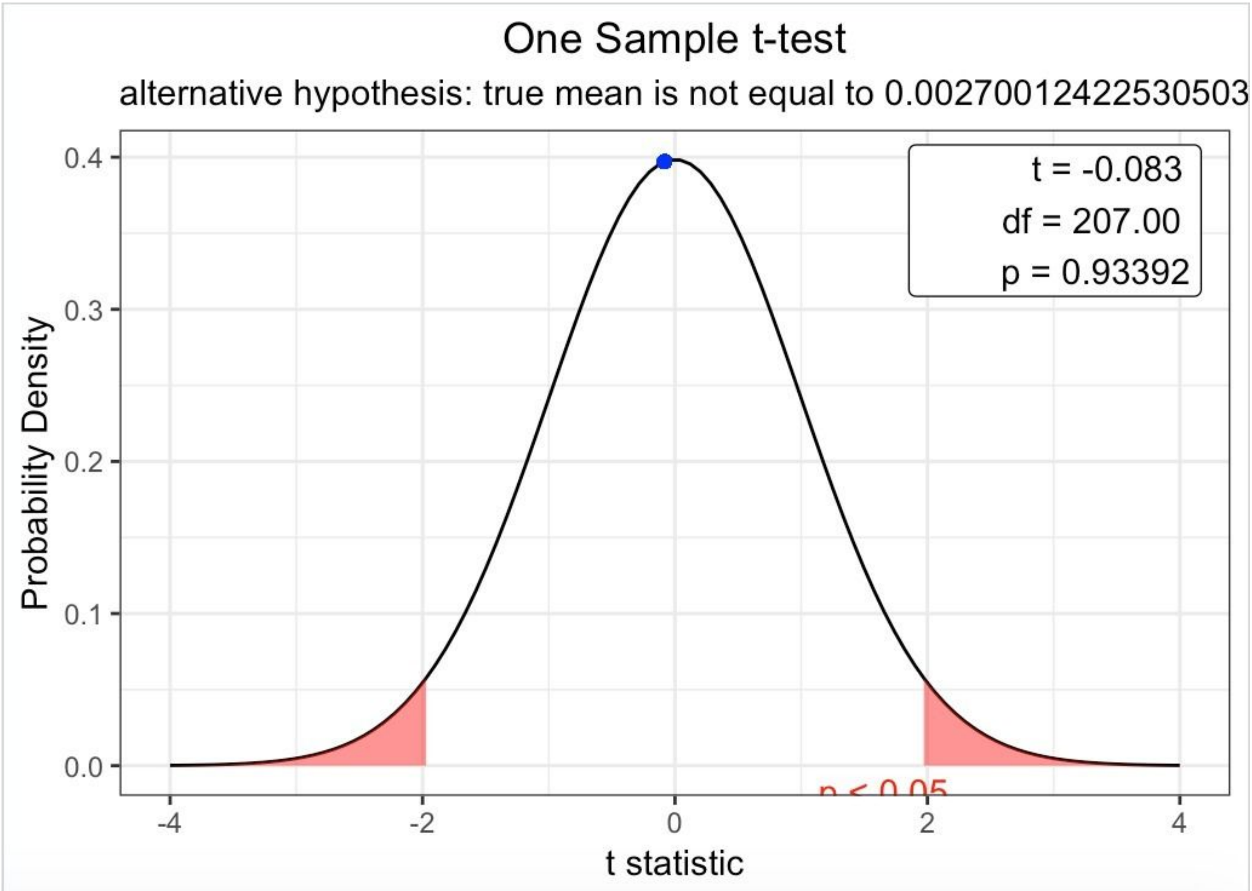
The T-Test of QQQ ETF is obtained as shown below:



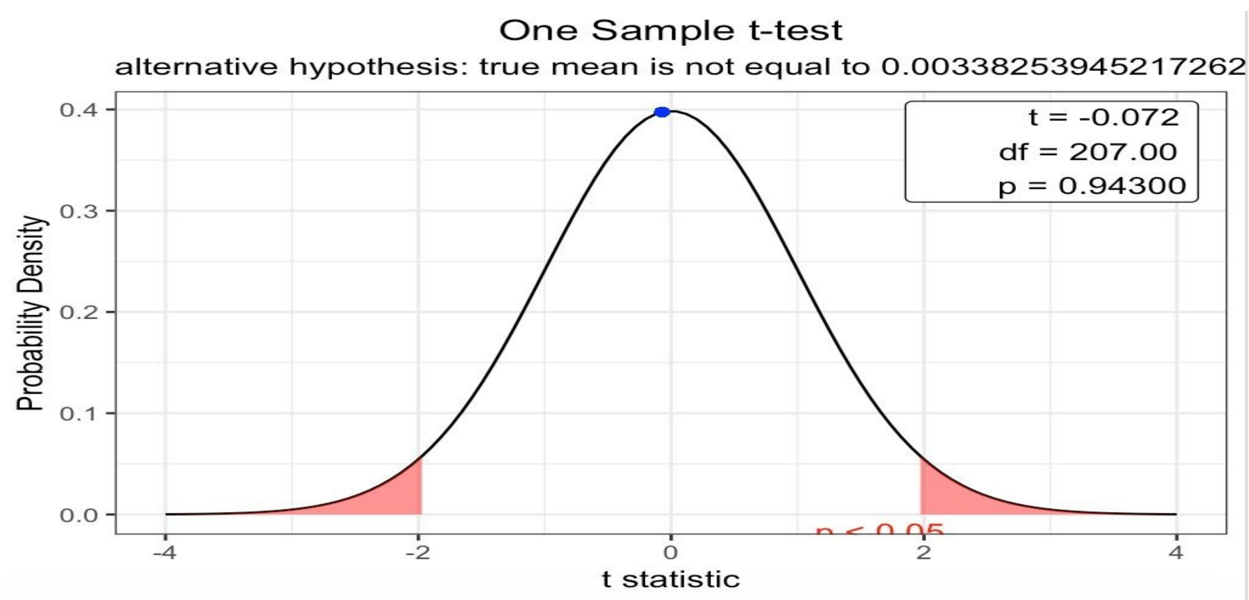

The T-Test of XLK ETF is obtained as shown below:



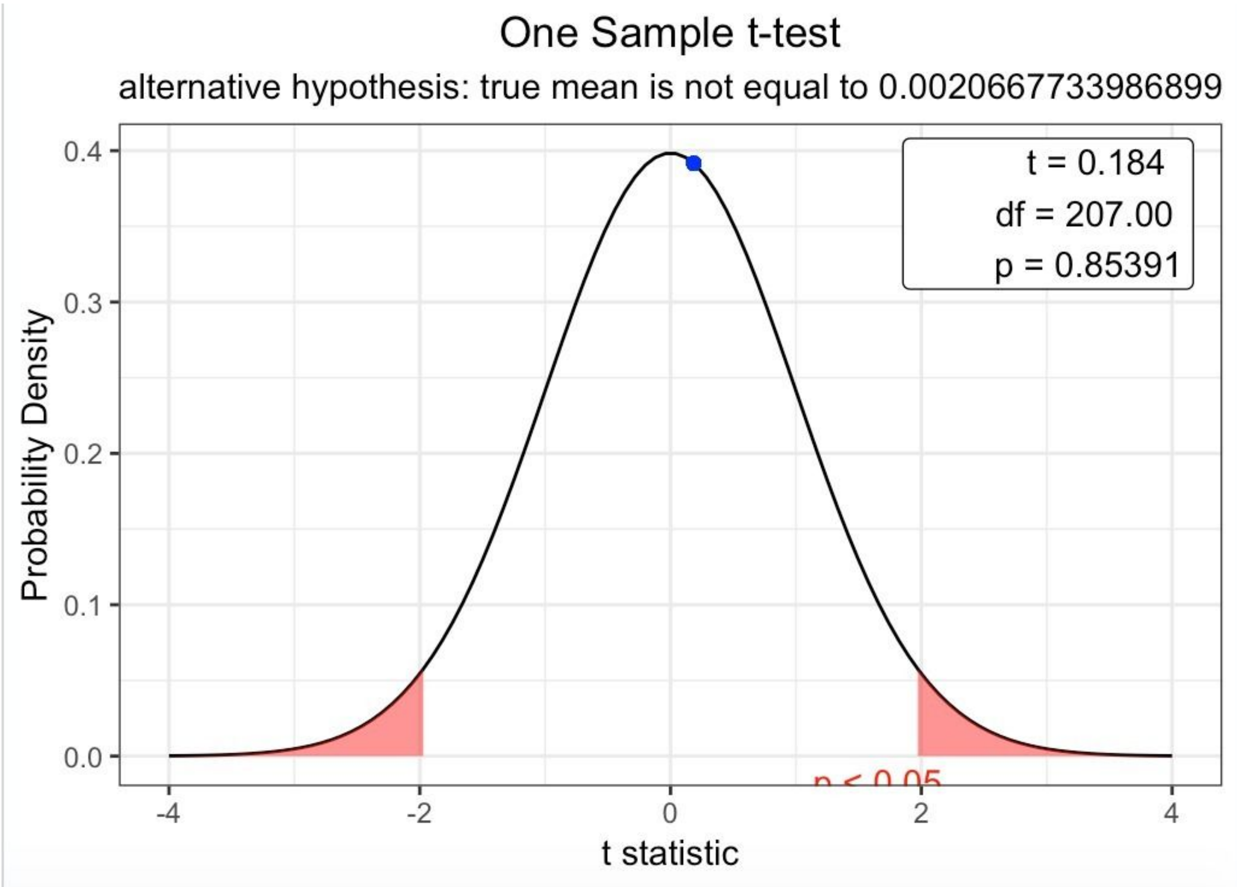

The T-Test Plot of PNQI ETF is obtained as shown below:




The T-Test Plot for FDN ETF is obtained as shown below:



The T-Test Plot for IYW ETF is obtained as shown below:




From all of these plots, we observe that the p-value ranges from 85% to 94%. This means that the null hypothesis more significant for each of the ETFs instead of the alternative hypothesis. Simply said, ETFs in this projects are very efficient.

## VI. Analysing Excess Returns of ETFs

The weekly returns of the observed five ETFs were calculated previously. The same approach was followed when calculating the weekly returns of the market indexes, which were the S&P 500 (S&P) and Dow Jones US Technology (DJUSTC). The returns were calculated using the log return formula.

On average, the weekly returns for the DJUSTC were 0.21%, which was 0.11% higher than that of the S&P 500. The Dow Jones US Technology Index invests in niche companies, particularly technology companies. On the other hand, S&P500's investments are spread across different industries and sectors, making it the basket of investments more diversified than the DJUSTC. By observing the standard deviation of the weekly returns of the DJUSTC and the S&P, we can see that the DJUSTC's weekly returns are higher due to higher risk taken from investment.

| | **Dow Jones** | **S&P 500** |
|---|---|---|
| Min | -8.594% | -6.923% |
| Median | 0.428% | 0.281% |
| Mean | 0.207% | 0.102% |
| Max | 6.108% | 3.968% |
| Standard Deviation | 0.02479 | 0.01701 |

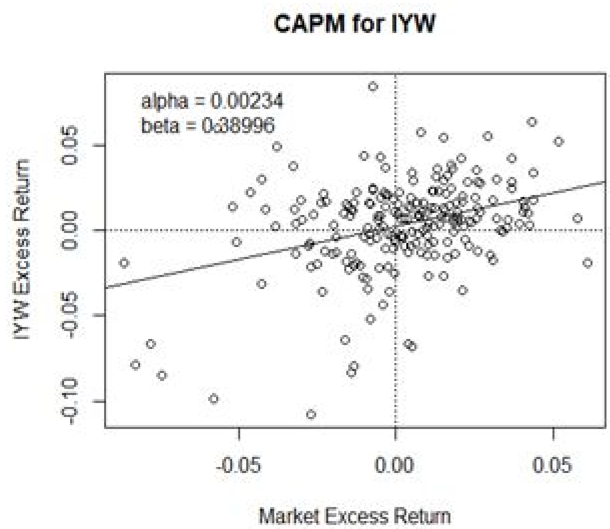
The excess return of an asset is the amount earned more than the risk-free rate with respect of time. The excess returns were calculated for each ETF and market indexes by subtracting the risk-free rate from weekly returns. When analyzing the results, positive excess returns indicate that the stock, or portfolio outperformed the riskless rate or benchmark.

On the other hand, negative excess returns occur when an investment underperforms in comparison to the risk-free rate or benchmark.

The Capital Asset Pricing Model (CAPM) is a model that estimates the return of an asset based on the return of the market. That asset's linear relationship to the return of the market is referred to as the stock's beta coefficient. The excess returns of the ETFs and market indexes were used to estimate the alphas and betas of the ETFs. Alpha and beta measure the risk and volatility of a security, or fund compared to that of a benchmark. Alpha observes the performance of the asset, while beta observes the level of inherent risk when comparing it to the Capital Asset Pricing Model. The table given below explains the regression on DJUSTC by each of the ETFs.

| | QQQ | XLK | PNQI | FDN | IYW |
|---|---|---|---|---|---|
| Alpha | 0.00132 | 0.00167 | 0.00151 | 0.00151 | 0.00234 |
| Beta | 0.3809 | 0.3732 | 0.3650 | 0.3669 | 0.39 |

The observations indicate that iShares U.S Technology ETF (IYW) had the highest alpha and beta. This tells us that the IYW's returns were more volatile than the other ETFs that were compared. In addition, the IYW's alpha shows that it outperformed the DJUSTC and is also greater than the other ETFs due to the higher level of risk that IYW assumed. Given below is a plot of IYW's excess returns with the market's excess returns.




VII. Evaluating the Model

Analysis of Variance (ANOVA) is a statistical method used to test the difference between means. There are three types of ANOVA tests. The one-way ANOVA test has one independent variable and one dependent variable. The Two-way ANOVA has one dependent variable and two independent variables. The N-way ANOVA observes more than two independent variables. The single-factor model of excess returns of an ETF regressed on the excess returns of a market index was observed by using the one-way ANOVA test. The test makes three general assumptions:

a. The population of samples tested are normally distributed.

b. Sample cases are independent of each other.

c. The variance among the groups should be approximately equal.

| ETFS | ANOVA ANALYSIS ON SINGLE FACTOR MODELS | | | | | |
|---|---|---|---|---|---|---|
| FDN | | Degree Freedom | Sums Squared | Mean Squared | F value | Pr(>F) |
| | DJUSTC Excess | 1 | 0.01713 | 0.01713 | 42.21123 | 6.04E-10 |
| | Residuals | 206 | 0.08358 | 0.00041 | | |
| IYW | | Degree Freedom | Sums Squared | Mean Squared | F value | Pr(>F) |
| | DJUSTC Excess | 1 | 0.01935 | 0.01935 | 28.17303 | 2.86E-07 |
| | Residuals | 206 | 0.14150 | 0.00069 | | |
| PNQI | | Degree Freedom | Sums Squared | Mean Squared | F value | Pr(>F) |
| | DJUSTC Excess | 1 | 0.01695 | 0.01695 | 36.49591 | 7.05E-09 |
| | Residuals | 206 | 0.09568 | 0.00046 | | |
| QQQ | | Degree Freedom | Sums Squared | Mean Squared | F value | Pr(>F) |
| | DJUSTC Excess | 1 | 0.01846 | 0.01846 | 44.33884 | 2.46E-10 |
| | Residuals | 206 | 0.08578 | 0.00042 | | |
| XLK | | Degree Freedom | Sums Squared | Mean Squared | F value | Pr(>F) |
| | DJUSTC Excess | 1 | 0.01772 | 0.01772 | 24.56833 | 1.49E-06 |
| | Residuals | 206 | 0.14859 | 0.00072 | | |

Based on this ANOVA table, the variability in the excess returns are significantly different for each ETF case. The probability of getting the F-values or a more extreme value, if the null hypothesis is true, is very small. Therefore, we reject the null hypothesis, which assumes that the means are equal.

| R-SQUARED ANALYSIS | | | | | |
|---|---|---|---|---|---|
| | QQQ | XLK | PNQI | FDN | IYW |
| **R-Squared** | 17.31% | 10.22% | 14.64% | 16.60% | 11.60% |

In addition, we observed the R-Squared values for each model to determine the good-fit. R-squared is a statistical measure of how close the data is  to the fitted regression line.

In general, higher the R-squared, the better the model fits your data. For instance, 100% indicates that the model explains all the variability of the response data around its mean. On the other hand, 0% indicates that the model explains none of the variability of the response data around its mean.

Based on the ratios, we see that they are below 20%. One observation that may cause a low percentage is the outliers in the data. Between December 2015 and January 2016, the Federal Reserve increased rates for the first time since they initiated Quantitative Easing. This increased volatility in the returns of equities for that period, including ETFs. This analysis tells us that it cannot describe the data appropriately. Hence, the model is not a good fit for the data that is used.

## VIII. Conclusion

We would like to thank Dr. Liping Ma for giving us this opportunity to analyse the performance of large AI ETFs and access their risk environments. Working on this project has taught a lot about Exchange-Traded Funds and what makes them an attractive investment opportunity.

We hope that the reader of this report was able to understand how to use the R programming language to perform various types of analysis on Exchange-Traded Funds and their indexes.

References

1. Staff, Investopedia (January 5, 2004). "Introduction to Exchange-Traded Funds". investopedia.com. Archived from the original on June 10, 2010.
2. "Exchange-Traded Funds (ETFs)". U.S. Securities and Exchange Commission. Archived from the original on November 11, 2013. Retrieved November 8, 2013.
3. Exchange-Traded Funds Archived July 6, 2017, at the Wayback Machine, SEC Release Nos. 33-8901, IC-28193, 73 Fed. Reg. 14618 (March 11, 2008).