SP500 Recent (2012-2022) Statistical Performance Analysis

Authors: Mahfuz Ibn Mannan, Sadula Srikrishna, Zhaojun He

Professor: Kenneth Abbott

WorldQuant University MScFE Capstone Project

E-mail: *mhfzmannan@gmail.com; krishna_sadula@yahoo.com; elizabeth.he777@gmail.com*

Abstract

We have statistically analyzed the weekly adjusted close prices and returns of four ETFs (U.S. SP500, VNQ, XLE, and XLF) from July 2012 to June 2022 (10 years historical rolling). The objective of this assignment is to revisit the performances of the ETFs considering the SP500 as the benchmark. We analyzed the historical data in the perspective of market dynamics and traced COVID-19 pandemic impact from regime shift identification. We checked the mean-reversion properties for covid 19 pandemic. As part of the analysis, stationarity, distributions, joint-distributions, associations, serial autocorrelation behavior, mean-reversion relationship, dependencies, etc. have been applied on four ETFs. Additionally, we have worked with sampled data (with Gaussian and non-Gaussian approach separately) besides the original data. This project is designed to consider the benchmark role of the SP500 and attempts to establish that the SP500 is a dependent ETF.

Keywords: *QQ-plot, ADF, CADF, PCA, k-Means, EDA, Multicollinearity, VIF, Stationarity, Serial Autocorrelation, Copulas, Markov Switching Model, Regime Shift Analysis, Mean-reversion*

1. Project Objective

Our main objective for this project is to conduct an extensive study on the SP500 ETF performances from July 2012 to June 2022. The plan is to consider the SP500 as a benchmark ETF and compare other three ETFs' significant movements in terms of distribution, joint distributions (with both original and transformed/sampled data), correlation, clustering, and stationarity. The SP500 ETF as a benchmark is more appealing to the financial market, as the index covers all the market ETFs and relatable parameters. We applied hyper-technical methods, Machine Learning (ML) tools in this project for regime shift identification and Covid 19 pandemic related mean-reversion property checks.

2.Literature Review

A great deal of research has been done with ETFs under many different scopes. The scopes of the works are diversified. Kajal and Moore (1991) worked on this scope of economic indicators [1]. They constructed leading economic indicators for the purpose of analyzing recessions. Underlying tests and visualizations are being carried out by Machine Learning approaches to track appropriate indicators. Another pattern of work is performance metrics of stocks. Sonam Srivastava, Mentor – Ritabrata Bhattacharyya [4] researched the instrument of the S&P 500 index from 2000-2017.

Furthermore, Mark Babayev, Folakemi Lotun, Googwill Tatenda Mumvenge, Ritabrata Bhattacharyya [5] did work emphasizing the black swan events. These black swan events might be important for a broad scale, especially for traders. Based on this, risk mitigation techniques can be applied if users have a clear understanding of trading performances. Hence, there is a possibility to construct an outperforming Mean-Reversion strategy.

The prediction performances of ETFs by Ligita Gaspareniene, Rita Remeikiene, Aleksejus Sodidko, and Vigita Vebraite [2] has increased to a remarkable level. In addition, Sonam Srivastava, Mentor – Ritabrata Bhattacharyya [4] conducted work on regime shift identification. We can use them for the appropriate time selection in investment, such as buying and selling of investments.

Gabjin Oh and Seunghwan Kim, Cheol-Jun Um (2006) worked for cross-border investors [6]. They addressed the statistical properties of daily index performance data (historical) of the S&P 500 with another ETF (KOSDAQ) of seven different countries. During the study, they used the Detrended Fluctuation and Surrogate tests. They found the returns of international stock market indices of those countries follow universal power laws.

In addition, emerging market ETFs (KOSDAQ) incurred a higher volatility than that of mature markets (S&P 500).

3.Background

Statistically, we revisited the S&P 500 performance of recent years. We aim to capture some rare events and their impact over the S&P 500. There are two points worth noting: First, is the symmetry, spreading, distributions, and inherited features. Second, the effect due to rare events. For instance, COVID-19 and the Russian attack on Ukraine. From a regime switching perspective, COVID-19 has the most significant impact for any market.

We chose to use adjusted closing price. It shows market sentiment of the trading day/week. An ETF is distinct from the S&P 500 index in many ways. Firstly, it weighs 80% of the stock market capital in the U.S. with the size of around 33.8 trillion USD in December 2022. Therefore, it reflects the U.S. economy to a reasonable extent. Our research focuses on the S&P 500 data from July 2012 to June 2022. We also considered three ETFs weekly returns in our research: Vanguard’s real estate index fund ETF (VNQ), an energy sector SPDR ETF (XLE), and a finance sector SPDR ETF (XLF).

4. Methodology

In this work, we applied some basic statistical measures, such as Exploratory Data Analysis and Unsupervised Machine Learning tools. Both basic statistical measures and EDA are simple and well recognized by researchers. We applied basic statistics, EDA and some UML algorithms (PCA, k-Means Clusters). Since we want to study the insights of data patterns, we used both PCA and K-means. We applied Markov Auto Regression Regime Switch Experiment to identify regimes. Furthermore, we used Cointegratted Augmented Dickey-Fuller (CADF) to analyze trend for covid-19 period. We also applied Copula (Gaussian and non-Gaussian) to check if the original dataset works appropriately and compare them to the sampled data found from Gaussian and Vine copulas.

4.1. Data Collection

We considered four ETFs in our work including Vanguard’s real estate index fund ETF (VNQ), an energy sector SPDR ETF (XLE), and a finance sector SPDR ETF (XLF), and the S&P 500 (SPY). We picked the S&P 500 as our benchmark since it is a stock market index tracker of the 500 largest companies listed in the U.S. These blue-chip companies represent 80% of the U.S. stock market capital, which embodies the U.S. economy to a significant degree. Our concentration was to evaluate the relationships from July 2012 to June 2022. We used python programming language through Jupyter notebook in this work. We used yfinance to extract relevant data from Yahoo Finance with the appropriate API Key.

4.2. Statistical Tests

We did some basic statistical tests including mean, median, standard deviation, skewness, and kurtosis. The results of these tests helped us to find the symmetric features and tailing pattern. We visualized the data through Exploratory Data Analysis (EDA), such as Box-plot (outlier detection), Scatter-plots and QQ-plot (normal distribution tracing). We conducted Jarque-Bera test to check sample data Skewness and Kurtosis in terms of normal distribution [3]. We attempted the Augmented Dickey-Fuller (ADF) test to analyze the stationarity of the series (first order serial autocorrelation). We calculated correlation coefficients to find out the degree and direction of association among each pair of the ETFs prices.

4.3 Hyper-technical Algorithm

4.3.1 PCA

Principal Component Analysis (PCA) is categorized as unsupervised learning, which is a well-established field of statistical approaches. This method has been used for quite some time and serves a variety of beneficial functions. The core concept is to reduce high-dimensional data to a much smaller set of dimensions so that the variables are independent. The methods we demonstrated in our project include comparing the covariance matrices of SPY, VNQ, XLE, and XLF. The PCA method takes high dimensional data and purifies it to its most crucial components. The Eigen-Decomposition of the matrix, which is the process of using historical data, returns, correlations, and covariances breaking down a matrix of covariances into pieces. The Sparse Principal Component Analysis (Sparse PCA) limits the number of variables by projecting into fewer dimensions. It uses only part of the stock rather than taking the conventional matrix and mapping it into a reduced space where entire variables are used.

In summary, the PCA is a form of unsupervised learning without labels. It is extremely useful in helping us understand data through Eigen-decomposition of the matrix. We are then able to reduce the data dimensions from a high-dimensional space to a lower-dimensional space. If we have a collection of independent variables, we can use eigenvalues to gauge the relative importance of each. Furthermore, it provides us a way to determine the relative importance of different dimensions in order to conduct further time series data analyses.

4.3.2 k-Means

Data is divided into clusters based on degree of similarity in the unsupervised machine learning technique known as data clustering. The most widely used method is K-Means clustering. There are five basic fundamental steps of K-Means, which can be listed as follows: In the first step, we decide how many clusters to search (In K-Means, this is the k). The algorithm then chooses k points at random from our data. These locations are also known as centroids. In the second step, each data point's distance from to centroids is determined. With the next step, each data point will be assigned to the closest centroid in order to create clusters. In the fourth step, a new centroid is determined by averaging the values of all its cluster members and discarding prior centroid values. During the final step, centroids are then calculated, again, by repeating steps two through four. The method is considered to converge when the centroids become stable.

Given enough data to train the algorithm, K-means may be used on any number of dimensions. However, in our project, we limit our use case with two dimensions in order to facilitate easy to read visualizations. The initial centroids chosen in step one are very critical in our project. For instance, even though it appears that the algorithm has discovered the optimal groupings, the fundamental steps were to be reinitiated with various initial centroids placements. In which case, a superior outcome might yet be discovered.

4.3.3 Markov Switching Model

The Markov Switching Autoregressive Model applies state-dependent parameters to dynamic regression models with varying characteristics throughout undetected states to account for structural breaks or multi-state occurrences. These models are referred to as Markov-switching models since a Markov chain is used to transition between the unobserved states. There are two types of models: Markov-switching autoregressive (MSAR) models ensuring regular adjustments, while regression (MSDR) models facilitate fast adjustments.

There are various models can be used to check the behavior of a series. The Markov switching model mainly emphasis on the behavior of the variables. Here are some parameters that are widely used:

1) Highest log likelihood: It brings the best fit of the model. Higher the value, the better the model.

2) AIC (Akaike information criterion) is a metric for determining fit of models. The lower the value of AIC, the better the fit of the model.

3) BIC (Bayesian Information Criterion) is a criterion for selecting a model from finite set of models.

4) HQIC (Hannan Quinn Information Criterion) is for model selection as well. It is an alternative to AIC & BIC.

We have tried 4 and 5 regimes, however, we decided to use 3 regimes to fit the model. 3 regimes are more evident and consistent.

4.3.4 Copula

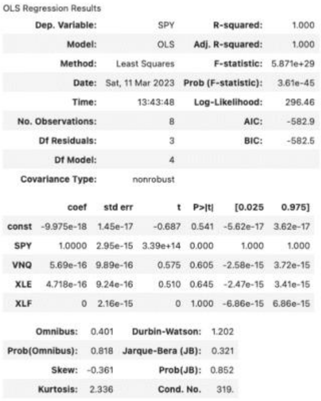
A copula is a mathematical function that connects distribution functions of more than one random variable to their respective marginal distribution functions. Therefore, it is a multivariate distribution function, and it is a way to trace dependencies.

5. Evaluation

We applied some traditional statistical measures and Exploratory Data Analysis (EDA) techniques, such as Box-plots, QQ-plots, joint-distribution plots for pair data series and Scatter Plots. Furthermore, applied Augmented Dicky-Fuller test (ADF) t-statistic to get a picture of stationarity (presence of unit root). We did first order serial autocorrelation.

Additionally, we applied correlation heat map and Variance Inflation Factor (VIF) to determine the multicollinearity. We used the Correlation matrix to calculate the degree and direction of association between ETFs prices. We also calculated Cointegrated ADF (CADF) to study the Mean-reversion property before and after the COVID-19 pandemic. We applied PCA and K-means to figure out the number of components that capture the variance. We especially paid attention to the COVID-19 impact. We applied the Markov Autoregressive Switching Model to identified the regime shift in the tenure of the study.

The evaluation methods we considered in this project are R-Squared, Akaike information criterion (AIC), Bayesian Information Criterion (BIC), Hannan Quinn Information Criterion (HQIC) and p-value. R-Squared is a statistical fit metric. A higher R-squared value indicates the more reliable the method is. The R-squared value listed in the figure below further confirms our model. The choice of AIC and BIC is based on HQIC. The p-value is the determining factor for consideration of unit root existence in our time series data.

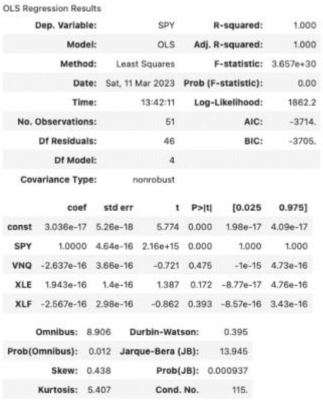


OLS Regression Results

Dep. Variable:	SPY	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	5.871e+29
Date:	Sat, 11 Mar 2023	Prob (F-statistic):	3.61e-45
Time:	13:43:48	Log-Likelihood:	296.46
No. Observations:	8	AIC:	-582.9
Df Residuals:	3	BIC:	-582.5
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-9.975e-18	1.45e-17	-0.687	0.541	-5.62e-17	3.62e-17
SPY	1.0000	2.95e-15	3.39e+14	0.000	1.000	1.000
VNQ	5.69e-16	9.89e-16	0.575	0.605	-2.58e-15	3.72e-15
XLE	4.718e-16	9.24e-16	0.510	0.645	-2.47e-15	3.41e-15
XLF	0	2.16e-15	0	1.000	-6.86e-15	6.86e-15

Omnibus:	0.401	Durbin-Watson:	1.202
Prob(Omnibus):	0.818	Jarque-Bera (JB):	0.321
Skew:	-0.361	Prob(JB):	0.852
Kurtosis:	2.336	Cond. No.	319.



OLS Regression Results

Dep. Variable:	SPY	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	3.657e+30
Date:	Sat, 11 Mar 2023	Prob (F-statistic):	0.00
Time:	13:42:11	Log-Likelihood:	1862.2
No. Observations:	51	AIC:	-3714.
Df Residuals:	46	BIC:	-3705.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.036e-17	5.26e-18	5.774	0.000	1.98e-17	4.09e-17
SPY	1.0000	4.64e-16	2.16e+15	0.000	1.000	1.000
VNQ	-2.637e-16	3.66e-16	-0.721	0.475	-1e-15	4.73e-16
XLE	1.943e-16	1.4e-16	1.387	0.172	-8.77e-17	4.76e-16
XLF	-2.567e-16	2.98e-16	-0.862	0.393	-8.57e-16	3.43e-16

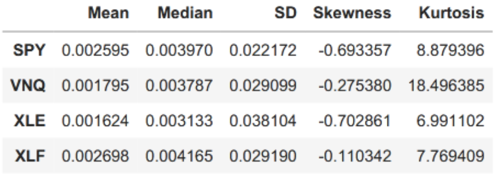
Omnibus:	8.906	Durbin-Watson:	0.395
Prob(Omnibus):	0.012	Jarque-Bera (JB):	13.945
Skew:	0.438	Prob(JB):	0.000937
Kurtosis:	5.407	Cond. No.	115.

6. Results

6.1 Basic Statistical Methods:

XLF looks paid highest, 0.27%, amongst all while XLE paid lowest, 0.18%, on average. In all ETFs median values are higher than that of respective mean values indicates non-normal distribution and left skewed (asymmetric distribution).

VNQ and XLF less skewed compared to other ETFs, indicates a nearly symmetric distribution. The greatest spreading is for XLE and the least for SPY around their respective mean. VNQ had a distinctively heavier tail amongst all return datasets. No kurtosis value was between -2 to +2, and hence all heavy tailed. As long as risky ETFs are concerned, XLE was riskier than all other ETFs since its SD value around mean was the highest.



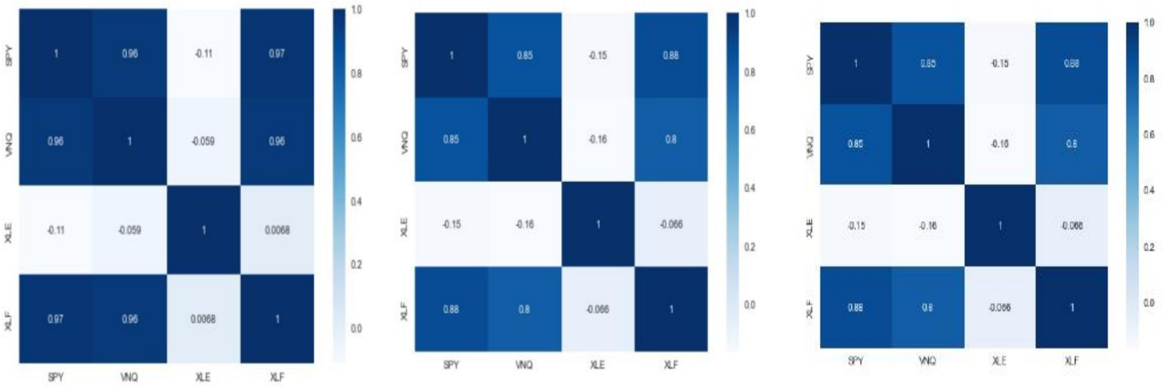
	Mean	Median	SD	Skewness	Kurtosis
SPY	0.002595	0.003970	0.022172	-0.693357	8.879396
VNQ	0.001795	0.003787	0.029099	-0.275380	18.496385
XLE	0.001624	0.003133	0.038104	-0.702861	6.991102
XLF	0.002698	0.004165	0.029190	-0.110342	7.769409

Table 1: Basic statistical parameters of the weekly returns

6.2 Correlation

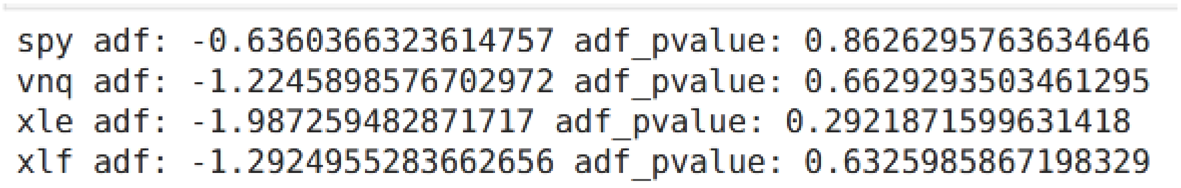
Correlation calculates pairwise linear relationship between ETFs. Both VNQ and XLF have significant correlation to SP500 which means strong associations of VNQ to SP500 and XLF to SP500 separately, while XLE looks negatively associated to SP500 with a lower degree. Hence, a combination of XLF with SP500 creates a diversified portfolio (in case of 2-ETF portfolio) and less risk tolerance. We calculated correlations in three ways: Spearman, Pearson and Kendall correlation matrix.

Table 2: Pearson (Left), Spearman (Middle), Kendall (Right)



6.3 Stationarity

Stationarity process is a distinct process where unconditional probability distribution remains constant over time. Both ADF t-statistic and KPSS tests are used to check stationarity of a time series. We applied ADF. The assessment process to check stationarity is to check p-values at 95% Confidence Interval. The null hypothesis states that there is a unit root in the time series. We could successfully be able to reject the null hypothesis since there is no p-value below 0.05 in our calculations. The lowest p-value is 0.29 for XLE. Therefore, our all-time series against all respective ETFs are with absence of unit root (not stationary).



```
spy adf: -0.6360366323614757 adf_pvalue: 0.8626295763634646
vnq adf: -1.2245898576702972 adf_pvalue: 0.6629293503461295
xle adf: -1.987259482871717 adf_pvalue: 0.2921871599631418
xlf adf: -1.2924955283662656 adf_pvalue: 0.6325985867198329
```

Table 3: p-values for prices of all 4 ETFs in ADF test

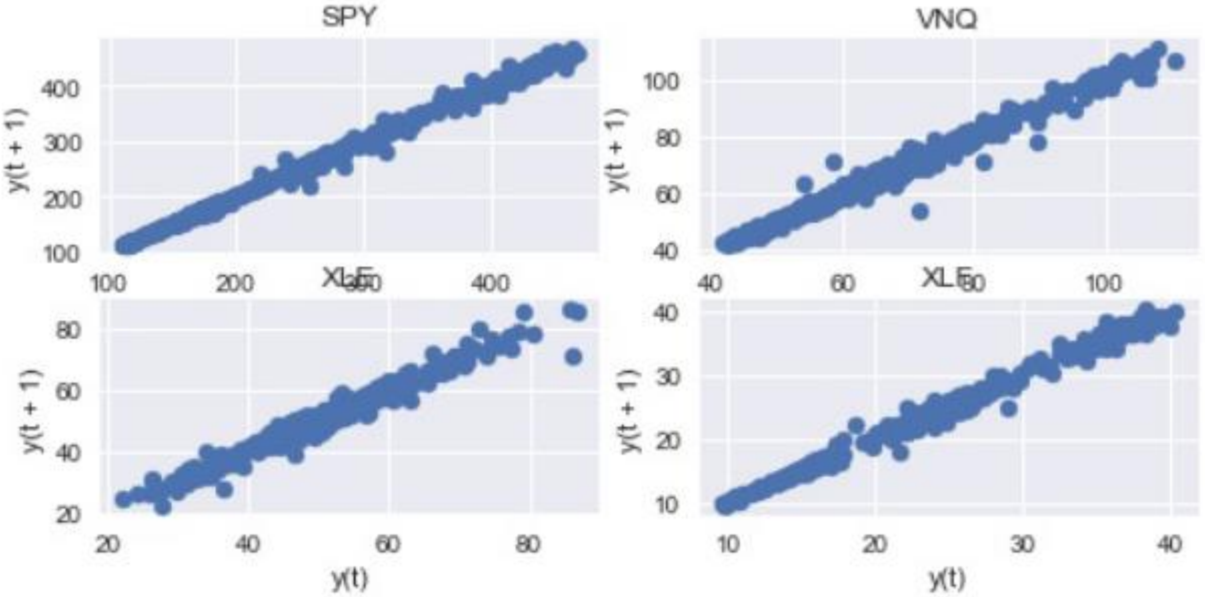


Table 4: First Order Serial Autocorrelation for all ETFs

If we consider the results from the above graphs with the outputs from QQ-plot, we will find that the returns don't follow the normal distributions. Any analysis assuming other than normal data distribution will be applicable to the return of instruments.

6.4 PCA

We applied four PCA components to study the method. Our findings are PCA-1 explained 73% of the variance in the data which is the most and PCA-2 explained 25%. Therefore, two PCA components are needed to express 98.33%. This way we got new set of un-correlated variables retaining most possible variations.

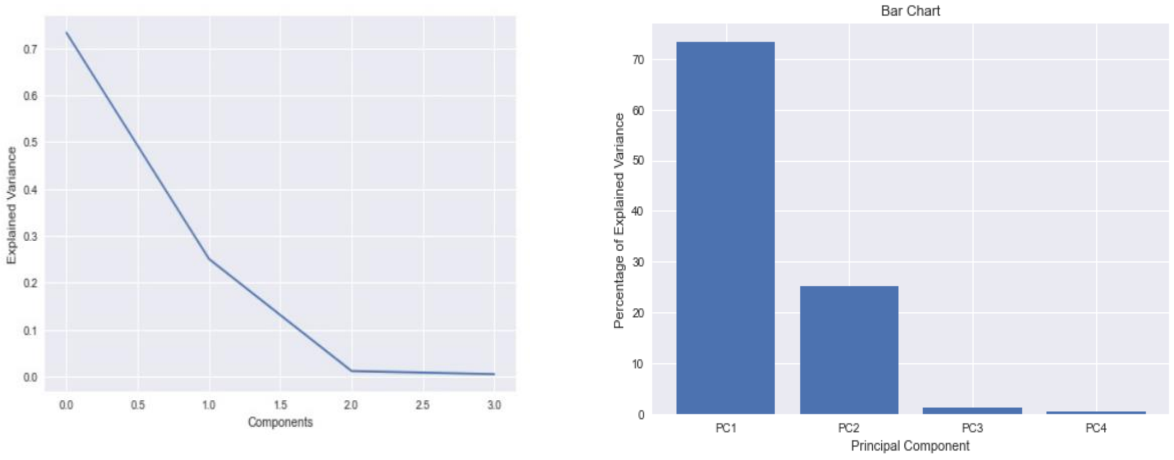
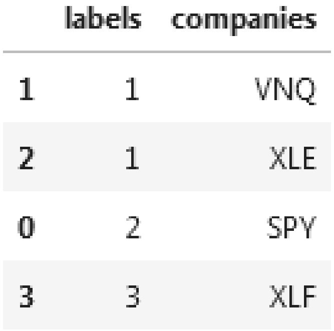


Table 5: PCA analysis table

6.5 k-Means

After we applied K-Means model, we found cluster 1 includes VNQ and XLE, whereas cluster 2 includes only SPY and cluster 3 includes XLF. Therefore, only VNQ and XLE are from the same group while SP500 and XLF are in the different and distinct groups. We have clustered our ETFs into 3 based on the degree of similarities in the project.



	labels	companies
1	1	VNQ
2	1	XLE
0	2	SPY
3	3	XLF

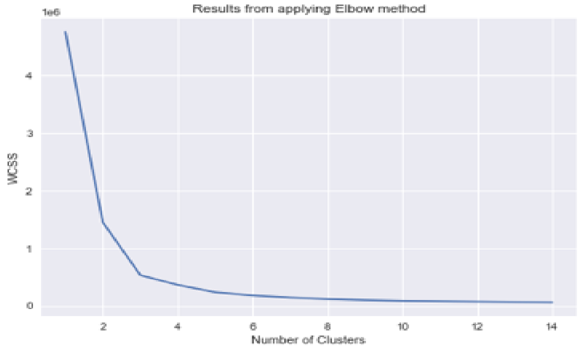
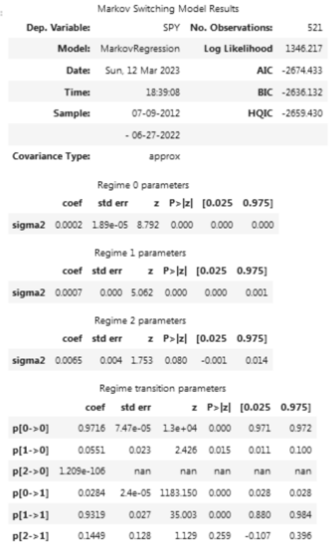


Table 6: k-Means Clustering (left hand side) and K-means Clustering (Elbow Method)

6.6 Markov Autoregression (MSAR)



Markov Switching Model Results

Dep. Variable:	SPY	**No. Observations:**	521
Model:	MarkovRegression	**Log Likelihood**	1346.217
Date:	Sun, 12 Mar 2023	**AIC**	-2674.433
Time:	18:39:08	**BIC**	-2636.132
Sample:	07-09-2012	**HQIC**	-2659.430
	- 06-27-2022		
Covariance Type:	approx		

Regime 0 parameters

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	0.0002	1.89e-05	8.792	0.000	0.000	0.000

Regime 1 parameters

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	0.0007	0.000	5.062	0.000	0.000	0.001

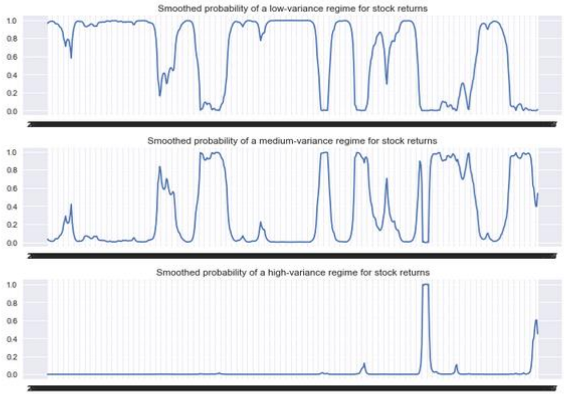
Regime 2 parameters

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	0.0065	0.004	1.753	0.080	-0.001	0.014

Regime transition parameters

	coef	std err	z	P>|z|	[0.025	0.975]
p[0->0]	0.9716	7.47e-05	1.3e+04	0.000	0.971	0.972
p[1->0]	0.0551	0.023	2.426	0.015	0.011	0.100
p[2->0]	1.209e-106	nan	nan	nan	nan	nan
p[0->1]	0.0284	2.4e-05	1183.150	0.000	0.028	0.028
p[1->1]	0.9319	0.027	35.003	0.000	0.880	0.984
p[2->1]	0.1449	0.128	1.129	0.259	-0.107	0.396

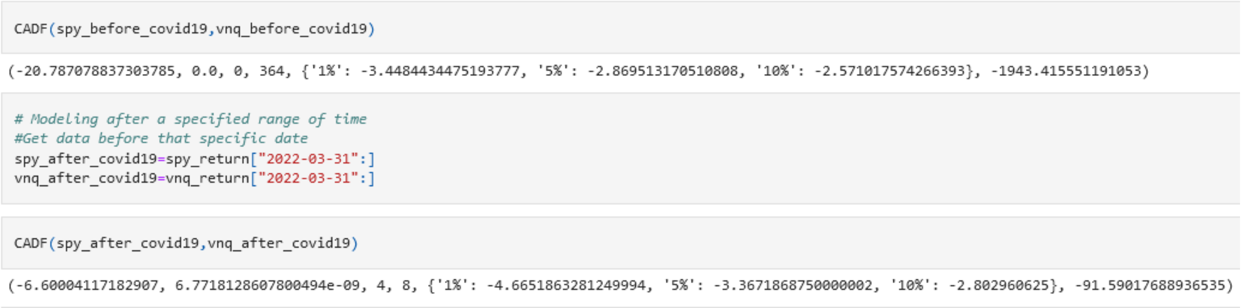
Table: For the SP500, the Markov Switching model results



Graph: Low, Medium, and High variance regimes

The assessment of MSAR is based on comparison between values of AICs that are found from several attempts of MSAR. These attempts are made with datasets of different time intervals. In this project, we have attempted MSAR only for a single dataset and the dates are observed based on the analysis of regime shift visualization graphs from Markov Auto Regression Regime Shift Experiment.

6.7 COVID-19



```
CADF(spy_before_covid19,vnq_before_covid19)
```

```
(-20.787078837303785, 0.0, 0, 364, {'1%': -3.4484434475193777, '5%': -2.869513170510808, '10%': -2.571017574266393}, -1943.415551191053)
```

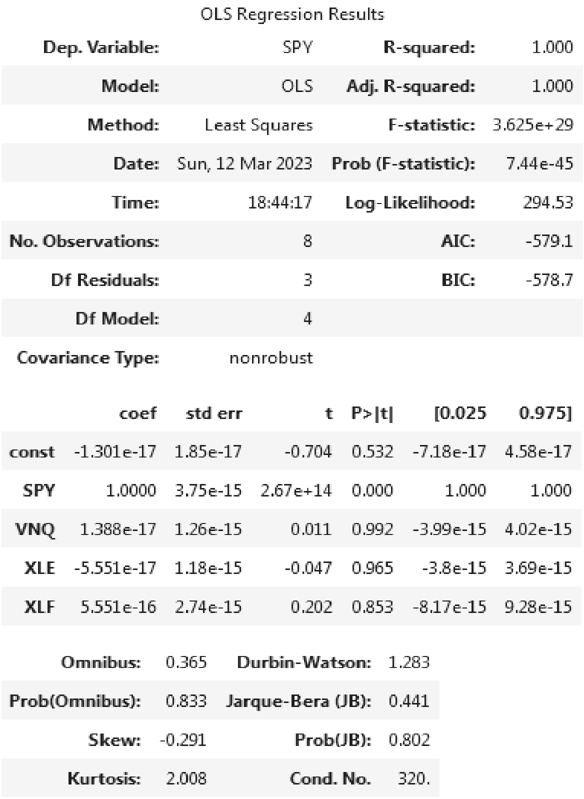
```
# Modeling after a specified range of time
#Get data before that specific date
spy_after_covid19=spy_return["2022-03-31":]
vnq_after_covid19=vnq_return["2022-03-31":]
```

```
CADF(spy_after_covid19,vnq_after_covid19)
```

```
(-6.60004117182907, 6.7718128607800494e-09, 4, 8, {'1%': -4.6651863281249994, '5%': -3.3671868750000002, '10%': -2.802960625}, -91.59017688936535)
```

CADF is applied to identify the mean-reversion relation for all four ETFs. The weekly returns of VNQ, XLE, and XLF are X variables while the SP500 is our proxy (benchmark that generates market movement). Our aim is to predict any relationship pairwise. Therefore, we applied multiple regression models. The entire dataset was split into two categories (pre-COVID-19 and post-COVID-19). Surprisingly, both the regression model results with maximum accuracy (R-Squared value of 1 for both pre and post COVID-19, which looks a bit abnormally high). These abnormalities suggested us to attempt the Correlation matrix and VIF for covid-19 to seek the multicollinearity.

From the values of the above picture the test statistics are -20.787 (for before COVID-19) and -6.600 (for after COVID-19). Both values are smaller than standard Critical values of 1%, 5% and 10%. We rejected the null hypothesis. CADF is used for identification of mean-reversion relation between datasets (here, we picked SP500 and VNQ). We found that there was a cointegrating relationship to a degree between these two ETFs.

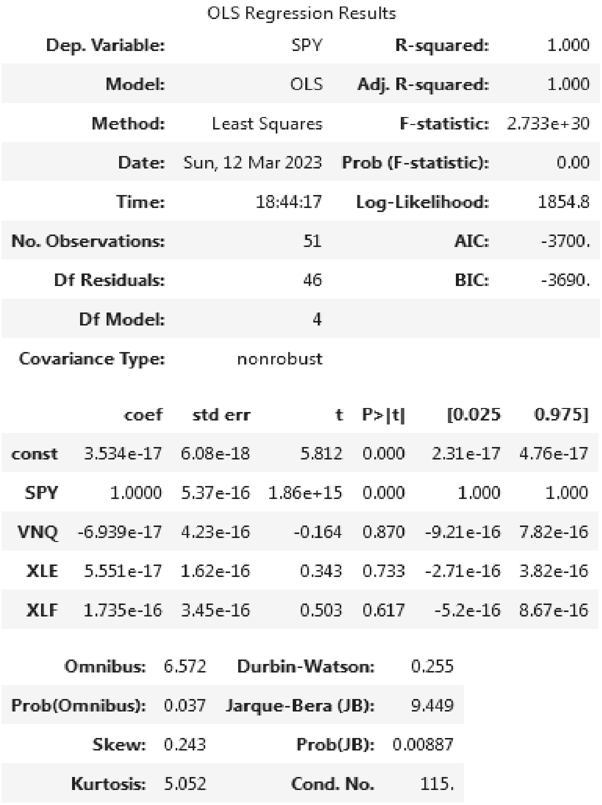


OLS Regression Results

Dep. Variable:	SPY	**R-squared:**	1.000
Model:	OLS	**Adj. R-squared:**	1.000
Method:	Least Squares	**F-statistic:**	3.625e+29
Date:	Sun, 12 Mar 2023	**Prob (F-statistic):**	7.44e-45
Time:	18:44:17	**Log-Likelihood:**	294.53
No. Observations:	8	**AIC:**	-579.1
Df Residuals:	3	**BIC:**	-578.7
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.301e-17	1.85e-17	-0.704	0.532	-7.18e-17	4.58e-17
SPY	1.0000	3.75e-15	2.67e+14	0.000	1.000	1.000
VNQ	1.388e-17	1.26e-15	0.011	0.992	-3.99e-15	4.02e-15
XLE	-5.551e-17	1.18e-15	-0.047	0.965	-3.8e-15	3.69e-15
XLF	5.551e-16	2.74e-15	0.202	0.853	-8.17e-15	9.28e-15

Omnibus:	0.365	**Durbin-Watson:**	1.283
Prob(Omnibus):	0.833	**Jarque-Bera (JB):**	0.441
Skew:	-0.291	**Prob(JB):**	0.802
Kurtosis:	2.008	**Cond. No.**	320.



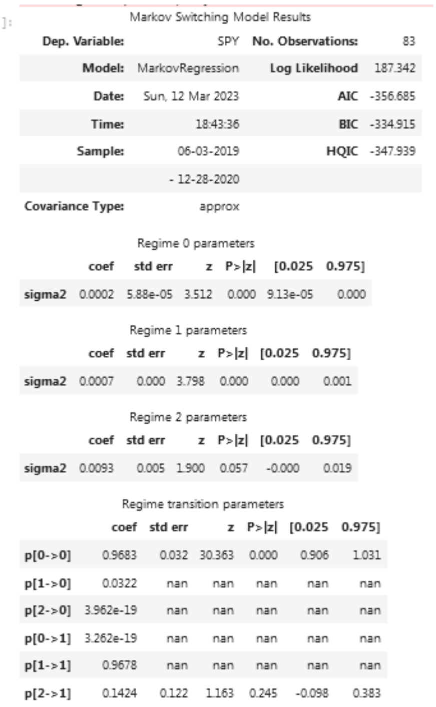
OLS Regression Results

Dep. Variable:	SPY	**R-squared:**	1.000
Model:	OLS	**Adj. R-squared:**	1.000
Method:	Least Squares	**F-statistic:**	2.733e+30
Date:	Sun, 12 Mar 2023	**Prob (F-statistic):**	0.00
Time:	18:44:17	**Log-Likelihood:**	1854.8
No. Observations:	51	**AIC:**	-3700.
Df Residuals:	46	**BIC:**	-3690.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.534e-17	6.08e-18	5.812	0.000	2.31e-17	4.76e-17
SPY	1.0000	5.37e-16	1.86e+15	0.000	1.000	1.000
VNQ	-6.939e-17	4.23e-16	-0.164	0.870	-9.21e-16	7.82e-16
XLE	5.551e-17	1.62e-16	0.343	0.733	-2.71e-16	3.82e-16
XLF	1.735e-16	3.45e-16	0.503	0.617	-5.2e-16	8.67e-16

Omnibus:	6.572	**Durbin-Watson:**	0.255
Prob(Omnibus):	0.037	**Jarque-Bera (JB):**	9.449
Skew:	0.243	**Prob(JB):**	0.00887
Kurtosis:	5.052	**Cond. No.**	115.

Table: Pre and Post Covid 19 (3 months before and after the Covid 19)



Markov Switching Model Results

Dep. Variable:	SPY	**No. Observations:**	83
Model:	MarkovRegression	**Log Likelihood**	187.342
Date:	Sun, 12 Mar 2023	**AIC**	-356.685
Time:	18:43:36	**BIC**	-334.915
Sample:	06-03-2019	**HQIC**	-347.939
	- 12-28-2020		
Covariance Type:	approx		

Regime 0 parameters

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	0.0002	5.88e-05	3.512	0.000	9.13e-05	0.000

Regime 1 parameters

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	0.0007	0.000	3.798	0.000	0.000	0.001

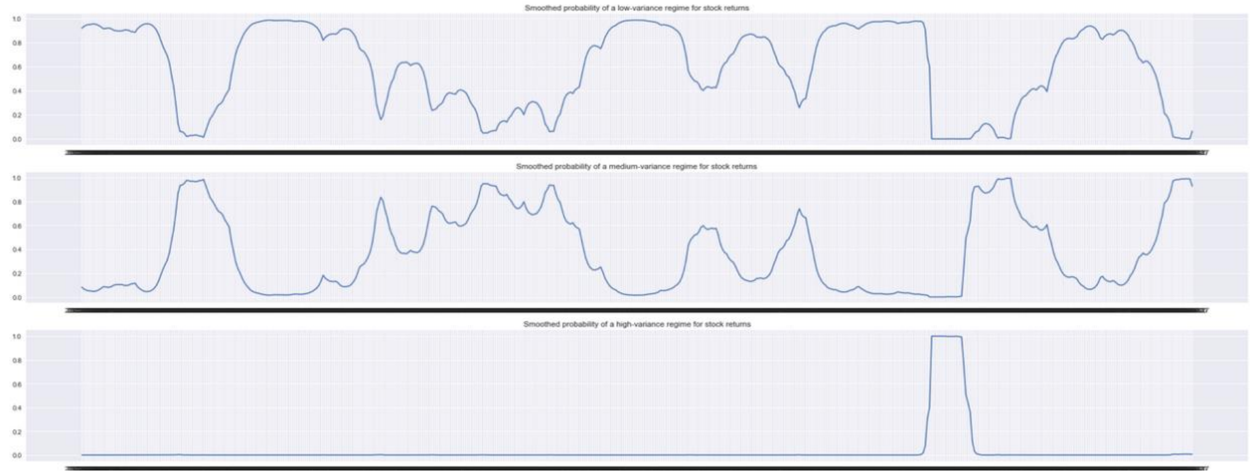
Regime 2 parameters

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	0.0093	0.005	1.900	0.057	-0.000	0.019

Regime transition parameters

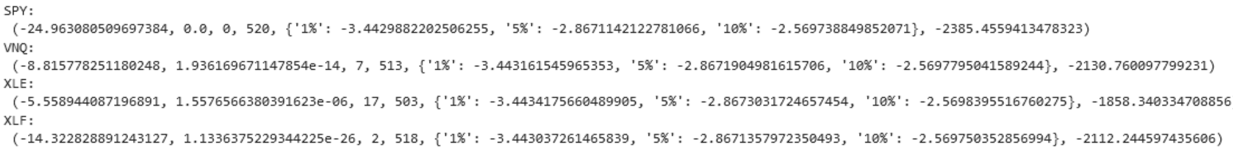
	coef	std err	z	P>|z|	[0.025	0.975]
p[0->0]	0.9683	0.032	30.363	0.000	0.906	1.031
p[1->0]	0.0322	nan	nan	nan	nan	nan
p[2->0]	3.962e-19	nan	nan	nan	nan	nan
p[0->1]	3.262e-19	nan	nan	nan	nan	nan
p[1->1]	0.9678	nan	nan	nan	nan	nan
p[2->1]	0.1424	0.122	1.163	0.245	-0.098	0.383

Table: Regime Shift Identification in COVID-19



Graph: Regime Shift test results

We applied regression analysis for both before and after COVID-19 and calculated the R-Squared value for both pre-COVID-19 and post-COVID-19. The values are 1 in both cases. This might be an amateur approach and there might be a misspecification in the model. Multicollinearity could be the main problem. Since we have very high VIF values for all four ETFs (except XLE) and the big correlation coefficients.



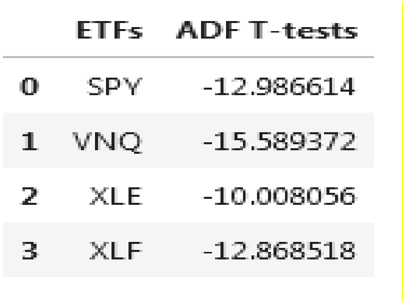
```
SPY:
 (-24.963080509697384, 0.0, 0, 520, {'1%': -3.4429882202506255, '5%': -2.8671142122781066, '10%': -2.569738849852071}, -2385.4559413478323)
VNQ:
 (-8.815778251180248, 1.936169671147854e-14, 7, 513, {'1%': -3.443161545965353, '5%': -2.8671904981615706, '10%': -2.5697795041589244}, -2130.760097799231)
XLE:
 (-5.558944087196891, 1.5576566380391623e-06, 17, 503, {'1%': -3.4434175660489905, '5%': -2.8673031724657454, '10%': -2.5698395516760275}, -1858.340334708856)
XLF:
 (-14.322828891243127, 1.1336375229344225e-26, 2, 518, {'1%': -3.443037261465839, '5%': -2.8671357972350493, '10%': -2.569750352856994}, -2112.244597435606)
```

Table: Regime Shift test results (ADF value)

The T-statistic values for all four ETFs were calculated. The test statistics are -24.963 (SP500), -8.816 (VNQ), -5.559 (XLE), and -14.323 (XLF). All these values are smaller than standard critical values of 1%, 5% and 10%. We rejected the null hypothesis. This implies there is less possibility of changes in course of time for every level for 1%, 5% and 10%.

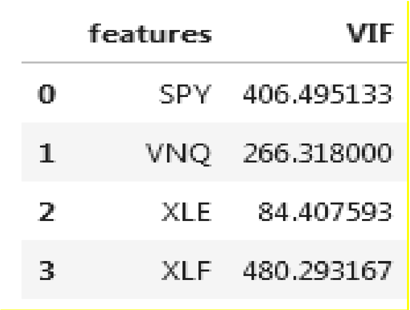
6.8 Mean Reversion

We applied ADF t-test for all the ETFs and found all negative integer values on the return time series, which indicates returns have mean-reverting properties. We found there were no trends that can be addressed as potential mean reverting properties. We did this with Cointegrated ADF. We concluded ETFs don't have any mean-reversion property, because statistically insignificant t-statistic value.



	ETFs	ADF T-tests
0	SPY	-12.986614
1	VNQ	-15.589372
2	XLE	-10.008056
3	XLF	-12.868518

Table: ADF T-statistic Values

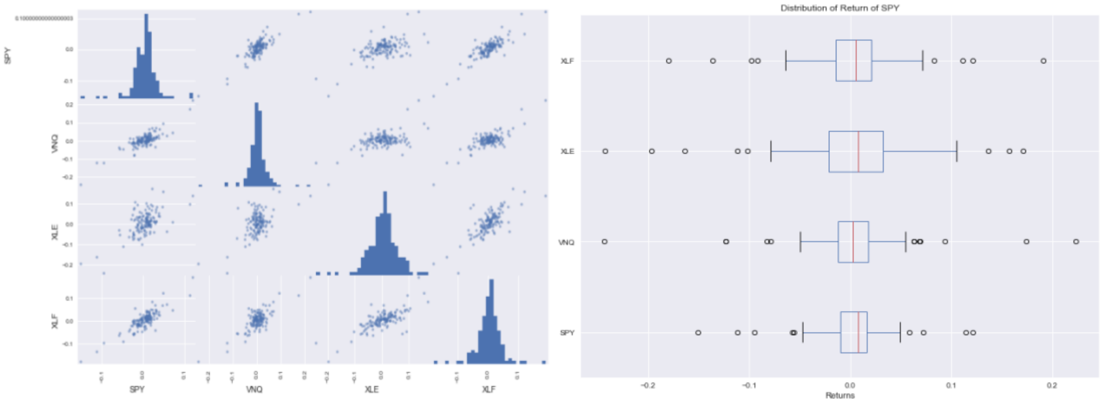


	features	VIF
0	SPY	406.495133
1	VNQ	266.318000
2	XLE	84.407593
3	XLF	480.293167

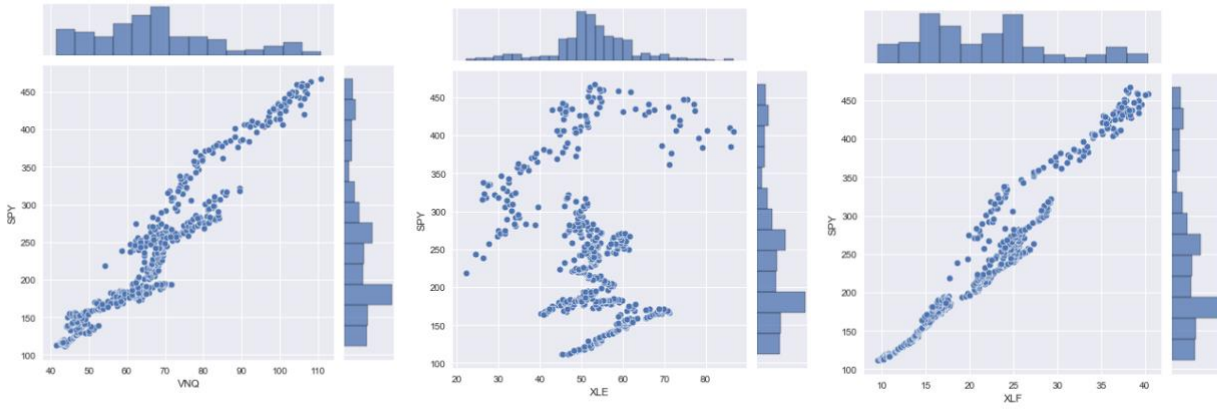
Table: VIF Values

6.9 Exploratory Data Analysis (EDA):

Box plots are used for EDA. It is an efficient way to identify and compare extreme values of any dataset. Boxplot displays distribution of data including IQR, Q1, median, Q3 and outliers. XLE indicates more volatile ETF during COVID-19 since it comprised of a widest box, whereas SP had small range of boxplot, which indicates the less risky feature of the ETF.

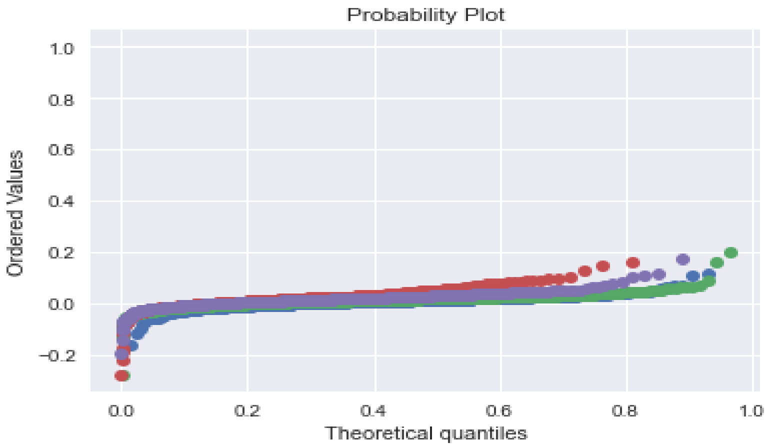


Graph: Scatter-plot (left hand side) and Box-plot (right hand side)

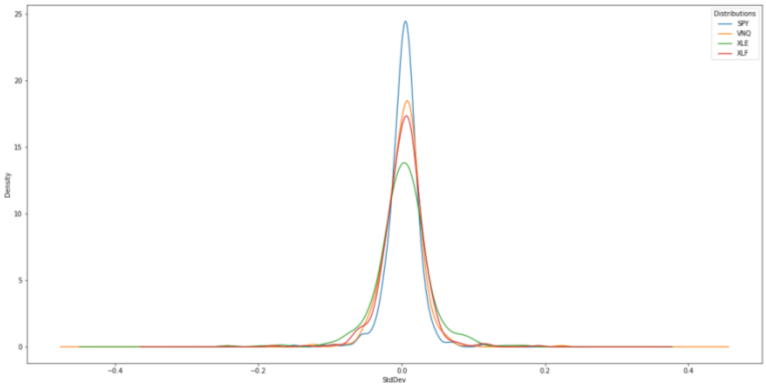


Graph: Joint distribution of prices of all ETFs against SP500

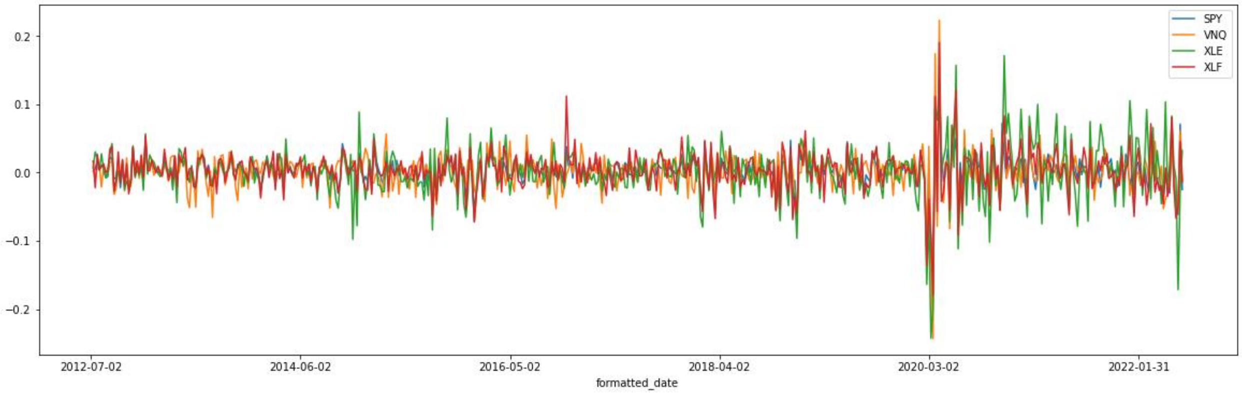
We drew a QQ-plot and calculated the p-value for the Jarque-Bera test. We found that the p-value is 0. We rejected null-hypothesis. This implies no data series distribution follows the normal distribution. We then applied the Density function as below.



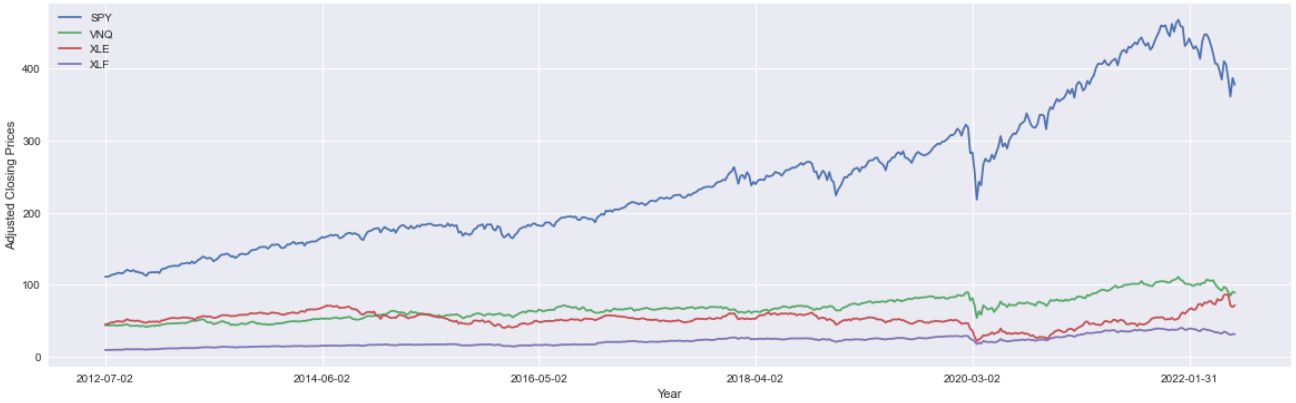
Graph: QQ-plot



Graph: Test for normality (Density function)



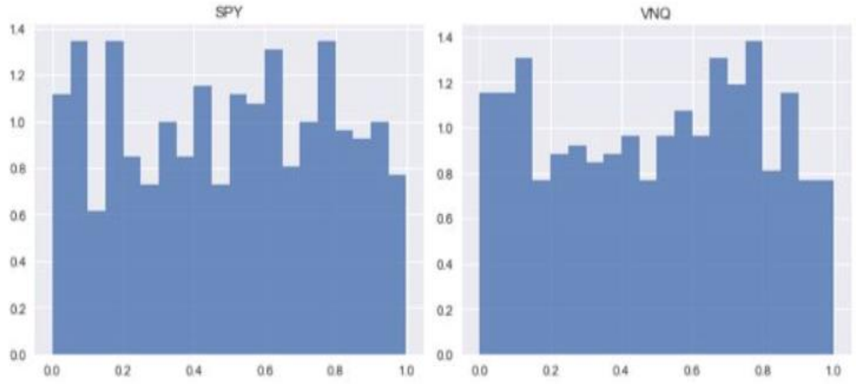
Graph: Return distributions



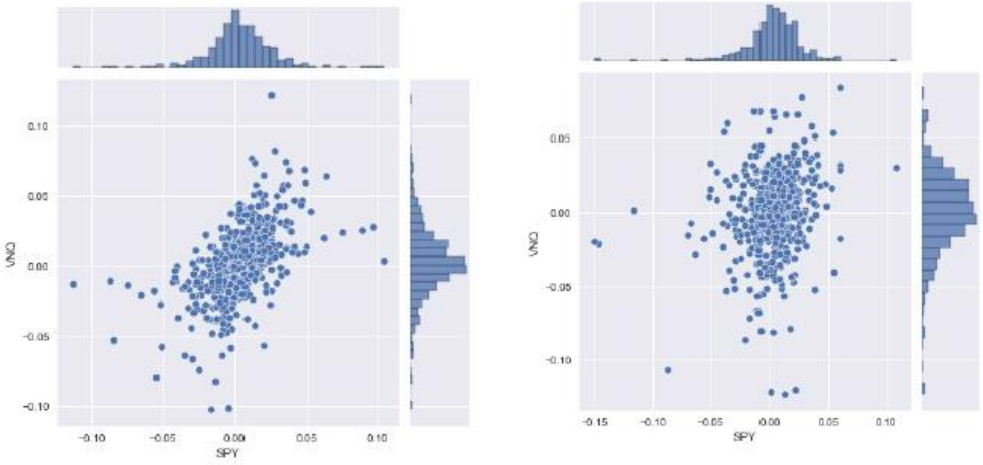
Graph: Adjusted Close Prices

6.10 Copulas

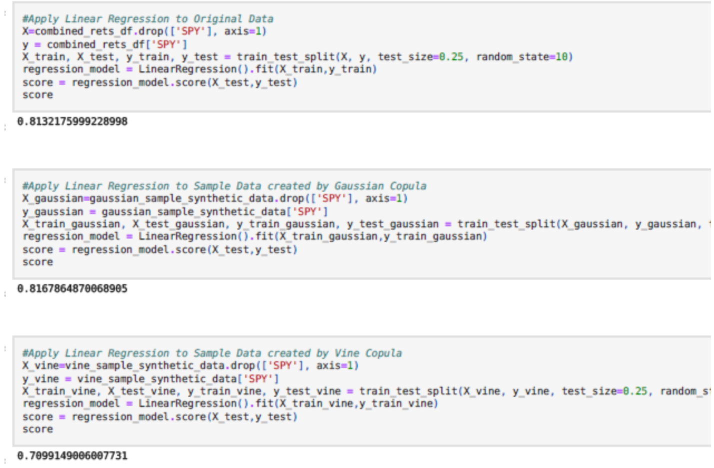
We have applied linear regressions for three distinct scenarios with original data.



Graph: Transformation of data to best available distribution by copula



Graph: Joint distribution of SPY and VNQ using transformed data using Gaussian Copula (left) and Vine Copula (right)



```
#Apply Linear Regression to Original Data
X=combined_rets_df.drop(['SPY'], axis=1)
y = combined_rets_df['SPY']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=10)
regression_model = LinearRegression().fit(X_train,y_train)
score = regression_model.score(X_test,y_test)
score
```

```
0.8132175999228998
```

```
#Apply Linear Regression to Sample Data created by Gaussian Copula
X_gaussian=gaussian_sample_synthetic_data.drop(['SPY'], axis=1)
y_gaussian = gaussian_sample_synthetic_data['SPY']
X_train_gaussian, X_test_gaussian, y_train_gaussian, y_test_gaussian = train_test_split(X_gaussian, y_gaussian,
regression_model = LinearRegression().fit(X_train_gaussian,y_train_gaussian)
score = regression_model.score(X_test,y_test)
score
```

```
0.8167864870068905
```

```
#Apply Linear Regression to Sample Data created by Vine Copula
X_vine=vine_sample_synthetic_data.drop(['SPY'], axis=1)
y_vine = vine_sample_synthetic_data['SPY']
X_train_vine, X_test_vine, y_train_vine, y_test_vine = train_test_split(X_vine, y_vine, test_size=0.25, random_s
regression_model = LinearRegression().fit(X_train_vine,y_train_vine)
score = regression_model.score(X_test,y_test)
score
```

```
0.7099149006007731
```

Figure: Linear Regression Scores with original data, sampled data (Gaussian and non-Gaussian approaches)

We have 3 scenarios: Scenario 1 is the application of linear model to the real dataset. We further have trained model with synthetic data (extracted from Gaussian and non-Gaussian copula models separately). Scenario 2 is the reflection of Gaussian copula fitting. Scenario 3 is related to non-Gaussian copula fitting. From the above picture, the scores from Linear, Gaussian, Non-Gaussian models are 0.8132, 0.8168 and 0.7099 respectively. The Gaussian model performance looks slightly more competitive than the original dataset.

6.11 Multicollinearity

We found correlation coefficients look high. There is a possibility of multicollinearity. Multicollinearity is one of the main problems of linear regression. Multicollinearity indicates the dependency among several independent variables (inter correlations between independent variables). We found both linear regression and Gaussian Copula have worked better than Vine copula.

7. Limitations

Our project utilized many Machine Learning approaches, such as PCA, k-Means, and Markov Switch Method. The method of our analysis can provide foundation for the prospective researchers.

In our project, we considered 3 ETFs with SP500. It would be better to consider more ETFs from different sectors (and industries) like power, pharmaceuticals, technology, insurance. We didn’t consider any macroeconomic factors in our study. Since ETFs’ performance is closely related to the macroeconomic factors, therefore further analysis can be conducted with macroeconomic indicators including fed rates, bank prime rates, unemployment rates.

8. Future work

In the future, works with alternative ETFs can be done with different time periods and different matrices. We can compare the results based on regimes to figure out trend and mean-reversion properties to detect any regulatory or deregulatory measures (quantitative easing, interest rate, fed rates, monetary policy, fiscal policy). We can apply different influential macroeconomic indicators (unemployment rate, bank loan rate, or commodity market index) over ETFs.

8. Conclusion

It is observed that four ETFs weekly returns are left skewed distributions with significant data residing at all the tail areas. Their serial autocorrelation looks nearly the same, which shows evidence that the SP500 is dependable on the other three ETFs and statistical properties remain unchanged for the other three ETFs. Correlations are strong. Variance Inflation Factor (VIF) for all the four is high. The joint analysis of correlation and VIF shows that one variable is dependent on all other three ETFs. We used cutting-edged technology, such as ML approaches and its application to study the regime switch and identified the regimes. Based on the regimes, we traced the Covid-19 timing. We studied the mean-reversion properties around Covid-19 and found no trend for Covid-19. We attempted to create sample datasets for covid 19 by fitting both Gaussian and vine copulas to our covid 19 related dataset and compared them with the original dataset.

Our code repo can be viewed through the following link:
https://github.com/krishxx/wqumscfin/blob/main/src/MScFE_Capstone.ipynb

References (MLA)

[1]. Lahiri, Kajal, and Geoffrey H. Moore, editors. "Leading Economic Indicators." *International Journal of Forecasting*, Cambridge UP, Jan. 1991

[2]. Ligita Gaspareniene, Rita Remeikiene, Aleksejus Sodidko, Vigita Vebraite, "Modeling of SP500 Index Price Based on U.S. Economic Indicators: Machine Learning Approach." *Inzinerine Ekonomika-Engineering Economics*, 2021, 32(4), p 362-375

[3]. Jarque-Bera test and its competitors for testing normality: A power comparison https://www.econstor.eu/bitstream/10419/49919/1/668828234.pdf

[4]. Sonam Srivastava, Mentor – Ritabrata Bhattacharyya, "Evaluating the Building Blocks of a Dynamically Adaptive Systematic Trading Strategy, *WorldQuant University*

[5]. Mark Babayev, Folakemi Lotun, Googwill Tatenda Mumvenge, Ritabrata Bhattacharyya, "Short Term Trading Models – Mean Reversion Trading Strategies and the Black Swan Events", *WorldQuant University*, SSRN

[6]. Gabjin Oh and Seunghwan Kim, Cheol-Jun Um, Statistical Properties of the Returns of Stock Prices of International Markets, arXiv: physics/0601126v1, 2006

Loretan, Mico, and William B. English. "Evaluating Correlation Breakdowns During Periods of Market Volatility." *SSRN Electronic Journal*, 2000. *Elsevier BV*, https://doi.org/10.2139/ssrn.231857.

Hamilton, J. D. (2005). Regime-Switching Models. *Palgrave Dictionary of Economics.*

Falbo, Paolo, and Rosanna Grassi. "Does Expectation of Correlation Breakdown In Financial Market Fulfill Itself?" Discrete Dynamics in Nature and Society, vol 2015, 2015, pp. 1-8. Hindawi Limited, https://doi.org/10.1155/2015/263908.