PROBLEM STATEMENT

MScFE Capstone Project

Mahfuz Ibn Mannan: mhfzmannan@gmail.com
Zhaojun He: elizabeth.he777@gmail.com
Sadula Srikrishna: krishna_sadula@yahoo.com

January 31, 2023

Project Track

For this Capstone project, we have selected a research track to analyze the topic of "Statistical Characteristics of Markets". A couple of market responsive events including regime changes, external shocks and alike expansion are under our purview. We will cover research on ETFs.

Problem statement

After the aftermath of the Global Financial Crisis markets started to rebound. Suddenly, Covid-19 pandemic appeared and amid the pandemic, another significant event started and that is the recent Ukraine War. It's a series of deadly events for the markets and therefore the market players (investors). In order to assess the investors' confidence over in recent decade we have planned to investigate couple of various ETFs including SPDR - S&P 500, Energy sector SPDR - XLE, Finance Sector SPDR - XLF, Vanguard Real Estate Index Fund ETF - VNQ for last decade ending 2022. We need to analyze with statistical sense to have a panorama of the consequences by the interventions of those rarely natural and man-made events and measures.

Yet many researches and reviews are obvious, nevertheless we understand this is an arena of continuous process of intense studies through which we are planning to do an extensive study weighing more to the recent events rather than far-fetching. Bearing this in mind we think to revisit and deep delve into these ETF performances along with relevant macroeconomic factors like Unemployment rates, Fed Fund rates, and Bank loan rates and then capture the footprint of the last decade from investors' lens.

Through this research any market player would have a latest and an indicative scenario in the ETF market. Therefore, assist the future investors to become proactive and current investors take measure with being informed about the latest studies.

Goals and Objectives

The main objective of this Capstone project is to revisit the performance of question ETFs over a decade through applying various traditional data analysis including Exploratory Data Analysis techniques and some ultra-modern Machine Learning tools along with some econometric models. Through taking a look at recent past history to explore each different market circumstances occurrence, we aim to do an occasional reassessment of a couple of ETFs

and identify any market regime and examine any external shock and explain in the light of relevant macroeconomic data analysis in statistical manner.

Decisions can be made in advance to account for any rare event like recession. Bankers and fund managers, for instance, may adopt conservative financial and investment strategies in reaction to an economic slowdown. Cutting back on loans and riskier investments while increasing cash reserves, central banks may pursue an expansionary monetary policy.

This Capstone project will be able to indicate to what extent historical consistent market regimes or conditions have existed. Countermeasures could be implemented by policymakers to lessen the severity of cycle downturns. The following statistical tests and data analysis frameworks are likely to be included based on relevance in this project.

(1) Exploratory Data Analysis: Exploratory data analysis is the crucial process of doing preliminary analyses on data in order to find patterns, identify anomalies, test hypotheses (Wrigley, 1979). For instance, we may apply box and whisker plots for extreme value identification. Furthermore, density plots, qq plots, Time series plots, conditional plots, Scatter plots, Tukey-mean difference plots are also our tools to apply based on necessity and relevance.

(2) Data Modeling: We may apply normal distribution to test the fitting data well. Alongside, a test for heavy-tailed distribution is also in our attention. Copula including non-Gaussian can be another tool in the modeling.

(3) Stationary Tests: We will divide our time series into partitions and compare the mean and variance. The null hypothesis that a unit root exists in a time series sample is commonly tested using the augmented Dickey-Fuller test (ADF) (Hadri & Larsson, 2005).

(4) Cointegration Check: To determine whether there is a long-term association between different time series, a cointegration test will be utilized. For instance, Error-Correction Model (ECM) is widely used for cointegrated data (Hendry & Juselius, 2001)

(5) Correlation: There is a chance that there will be some correlation, because time series data are continuous and chronologically ordered. The association of multicollinearity is problematic (O'Hagan & McCabe, 1975). When fitting the model and interpreting the findings, it may be difficult if there is a high enough degree of correlation between the variables.

(6) Return and Volatility: Extreme volatility and returns have a significant impact on investor sentiment in the stock market. When investors meet exceptionally pessimistic circumstances, positive or negative returns can quickly increase or diminish the level of investor sentiment (Gong, et al. 2016). As the mood lifts, the influence steadily lessens until the situation becomes upbeat.

(7) Regime Shift: There is an unequal relationship between the stability of monetary policy regimes and the volatility of stock prices (Ivrendi & Guloglu, 2012). Regimes can be established in a variety of ways. We will use several machine learning models learned in the previous courses to analyze and detect regime shifts based on market experience. In the case of regime shift, we may apply some econometric models like ARMA, GARCH, Cointegration and etc (where applicable and relevant) before and after the regime shift.

(8) Hyper-technical tools: Couple of Machine Learning (ML) techniques can be applied. LASSO, PCA, CART, SVM, NN might be needed in our assessment especially for pre and post regime shift data analysis. We may work with mean-reverting relationships to this extent. For model selection purposes it is helpful to study Cluster Analysis, Tree etc.

(9) Others: Extreme value determination can be one of the addressing points and in that tune standard deviation, skewness, kurtosis, trimmed-mean can be applied to capture any outliers.

Code Design / Pseudo Code

Following steps will be taken up as part of the code design.

1. Sourcing data from Yahoo Financial & Fred economic data
2. SP500, 3 different sectors ETFs & rates data are being chosen for this analysis.
3. Carryout the EDA (Exploratory data analysis) on the chosen financial data
4. Finding out statistical characteristics of the data
5. Identifying the regime shifts

Sample screenshots are added here, the corresponding code is being update at the github location: https://github.com/krishxx/wqumscfin/blob/main/src/MScFE_Capstone.ipynb

jupyter MScFE_Capstone Last Checkpoint: 4 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Data Sourcing

Obtain the data from 1 Oct 2012 to 30 September 2022 for the following: (say) 2012-10-01 to 2022-09-30

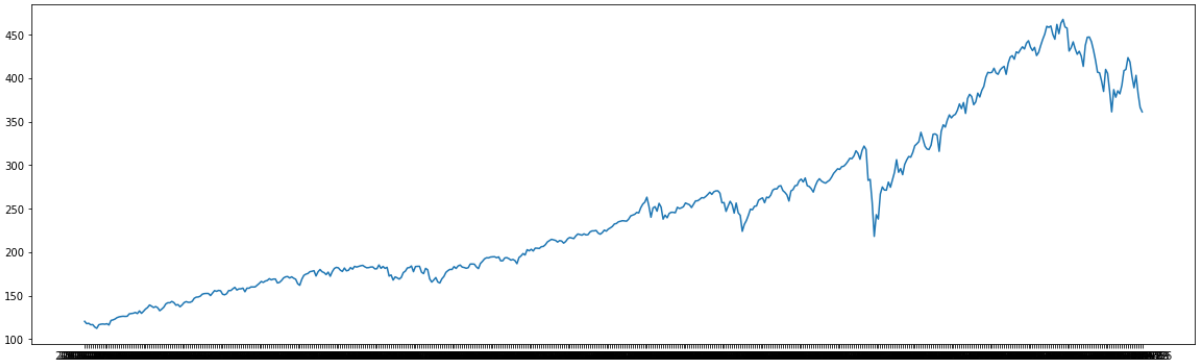
1. SPY (SPDR S & P 500 ETF)
2. XLE (Energy Sector SPDR)
3. VNQ (Real Estate)
4. XLF (Finance Sector SPDR)

In [24]:
```
import pandas as pd
import yfinance as yf
from yahoofinancials import YahooFinancials
```

In [25]:
```
yahoo_financials = YahooFinancials('SPY')
data = yahoo_financials.get_historical_price_data(start_date='2012-10-01',
                                                  end_date='2022-09-30',
                                                  time_interval='weekly')
spy_df = pd.DataFrame(data['SPY']['prices'])
spy_df = spy_df.drop('date', axis=1).set_index('formatted_date')
spy_df.head()
```

Out[25]:

formatted_date	high	low	open	close	volume	adjclose
2012-10-01	147.160004	143.830002	144.520004	146.139999	619770200	120.519783
2012-10-08	146.119995	142.580002	145.600006	142.889999	599319200	117.839569
2012-10-15	146.520004	142.770004	143.229996	143.389999	679092400	118.251930
2012-10-22	143.669998	140.389999	143.149994	141.350006	718295200	116.569557
2012-10-29	143.720001	140.679993	141.850006	141.559998	342136300	116.742752

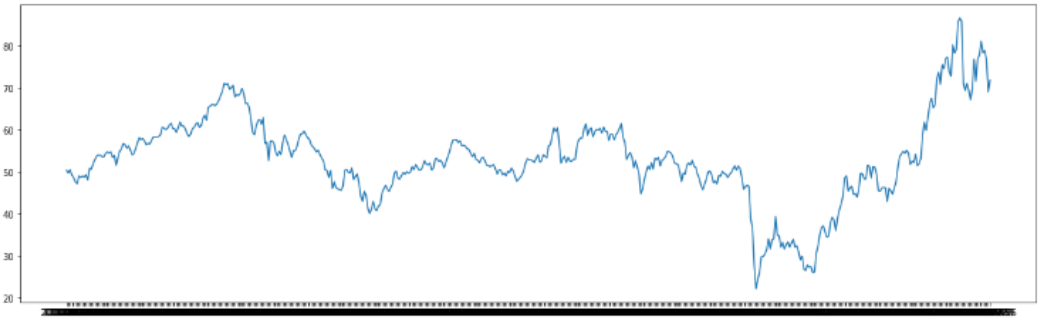


In [26]:
```
yahoo_financials = YahooFinancials('XLE')
data = yahoo_financials.get_historical_price_data(start_date='2012-10-01',
                                                  end_date='2022-09-30',
                                                  time_interval='weekly')
xle_df = pd.DataFrame(data['XLE']['prices'])
xle_df = xle_df.drop('date', axis=1).set_index('formatted_date')
xle_df.head()
```

Out[26]:

formatted_date	high	low	open	close	volume	adjclose
2012-10-01	74.529999	72.669998	73.940002	73.580002	57754800	50.376877
2012-10-08	74.379997	72.040001	73.169998	72.559998	49958700	49.678543
2012-10-15	75.190002	71.910004	72.430000	73.830002	60294000	50.548058
2012-10-22	74.099998	71.019997	73.860001	71.959999	57769500	49.267746
2012-10-29	72.690002	70.820000	72.250000	71.059998	31881500	48.651554

In [51]:
```
fig, ax = plt.subplots(figsize=(20,6));
ax.plot(xle_df['adjclose']);
```



Reference (MLA)

Wrigley, Neil. "Exploratory Data Analysis." *Area*, vol. 11, no. 1, 1979, pp. 18–19.

Hadri, Kaddour, and Rolf Larsson. "Testing for Stationarity in Heterogeneous Panel Data Where the Time Dimension Is Finite." *The Econometrics Journal*, vol. 8, no. 1, 2005, pp. 55–69.

Hendry, David F., and Katarina Juselius. "Explaining Cointegration Analysis: Part II." *The Energy Journal*, vol. 22, no. 1, 2001, pp. 75–120.

O'Hagan, John, and Brendan McCabe. "Tests for the Severity of Multicolinearity in Regression Analysis: A Comment." *The Review of Economics and Statistics*, vol. 57, no. 3, 1975, pp. 368–70.

Gong, Xu, et al. "Extreme Return, Extreme Volatility and Investor Sentiment." *Filomat*, vol. 30, no. 15, 2016, pp. 3949–61.

Ivrendi, Mehmet, and Bulent Guloglu. "Changes in Stock Price Volatility and Monetary Policy Regimes: Evidence from Asian Countries." *Emerging Markets Finance & Trade*, vol. 48, 2012, pp. 54–70