THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Analysis of USDX and ETFs

SEAS 6401
Instructor: Benjamin Harvey
Group Member:
Bite Xiong
Qing Zhu

Outlines

- Introduction
- Exploratory Data Analysis
- Forecasting
- Conclusion

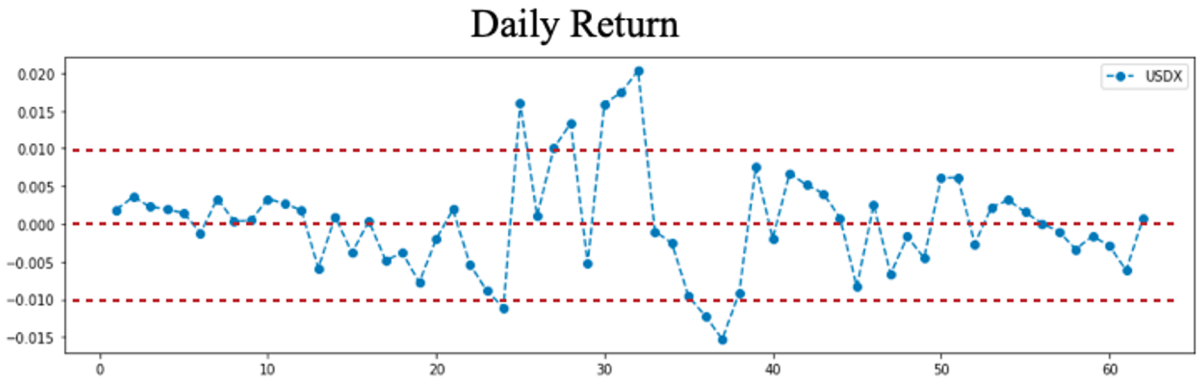
Introduction

- Covid-19 Pandemic
- Stock Market Downward
- Explore a Strategy
- USDX & ETFs
  02/01/2020-05/01/2020

Exploratory Data Analysis

Return Analysis

USDX: Index Tracing U.S Dollar
Total Return: 2%, The **Only** positive one
Daily Return: Slightly up&down



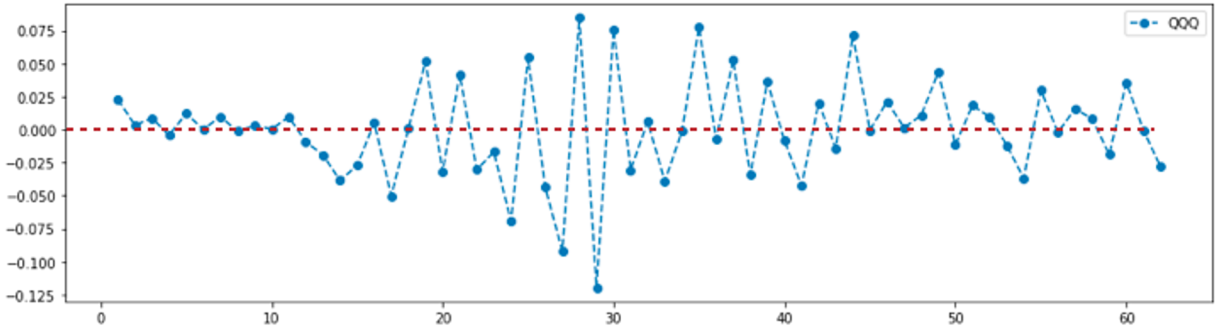
Exploratory Data Analysis

Return Analysis

QQQ: Nasdaq-100 (Apple, Microsoft etc.)
Total Return: -4%
Daily Return: Maximum **12.5%** down

Total Return

Daily Return



Exploratory Data Analysis

Return Analysis

XLE: Energy
Total Return: -32%
Daily Return: More Points below 0 scale

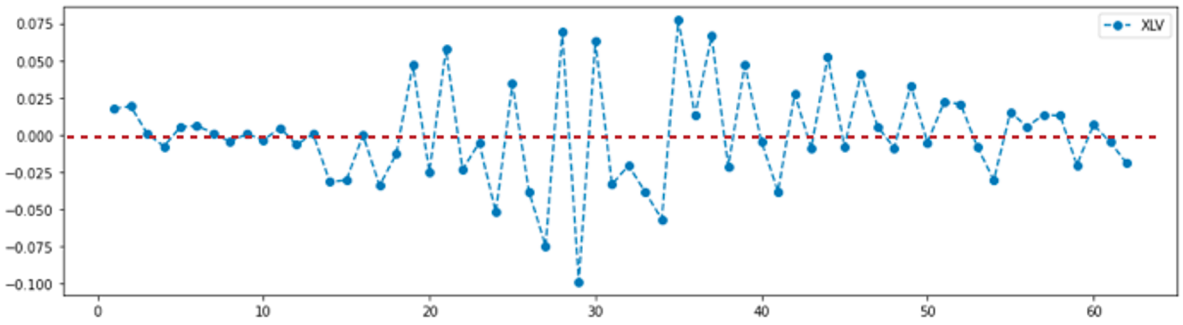
Exploratory Data Analysis

Return Analysis

XLV: Health care
Total Return: -2%
Daily Return: Symmetry Distributed

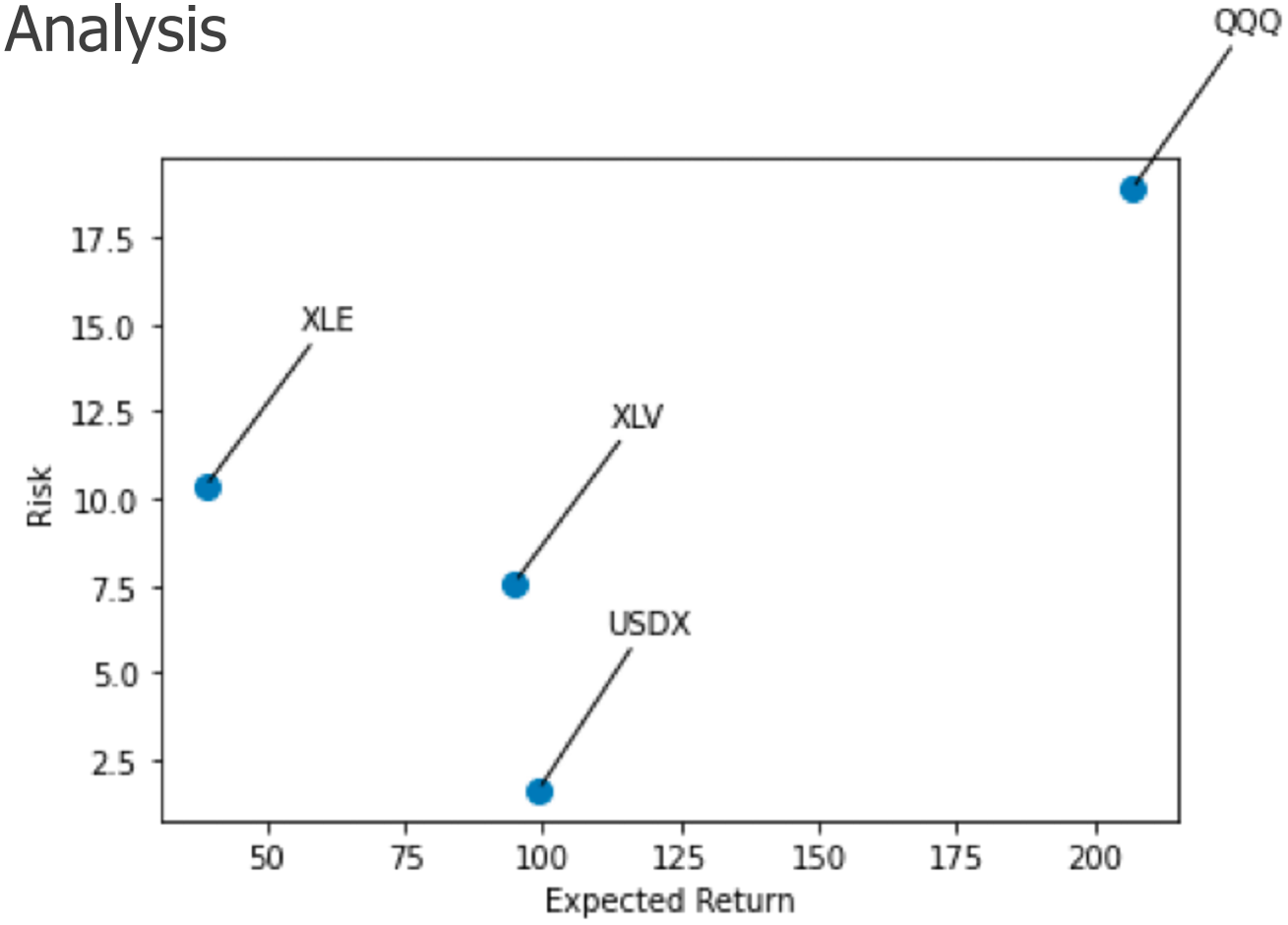
Total Return

Daily Return



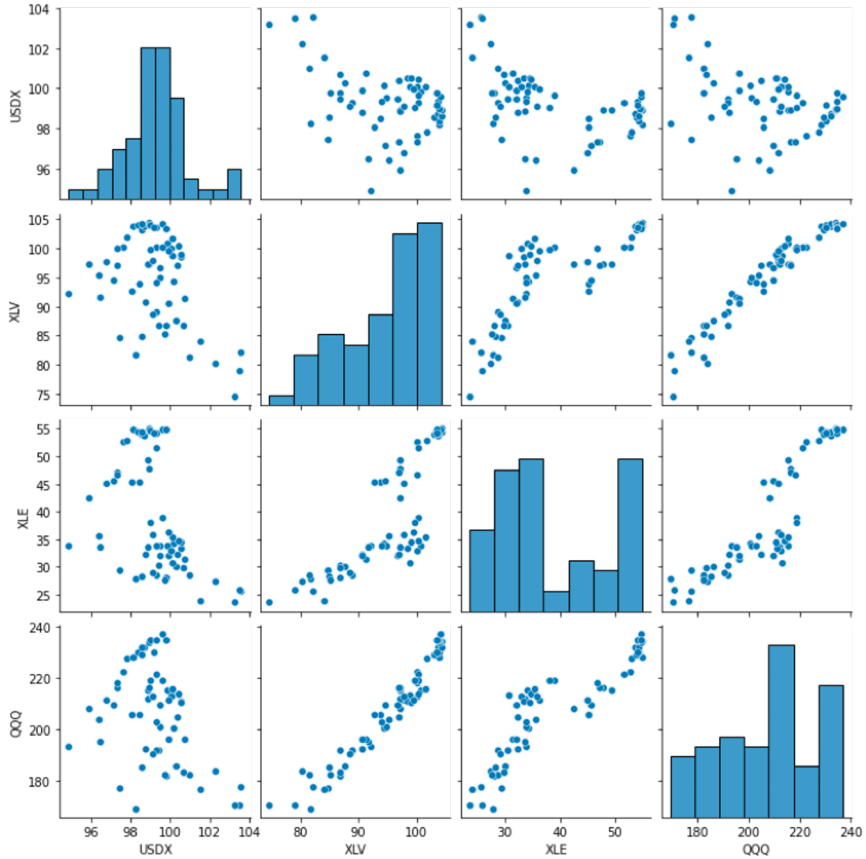
Exploratory Data Analysis

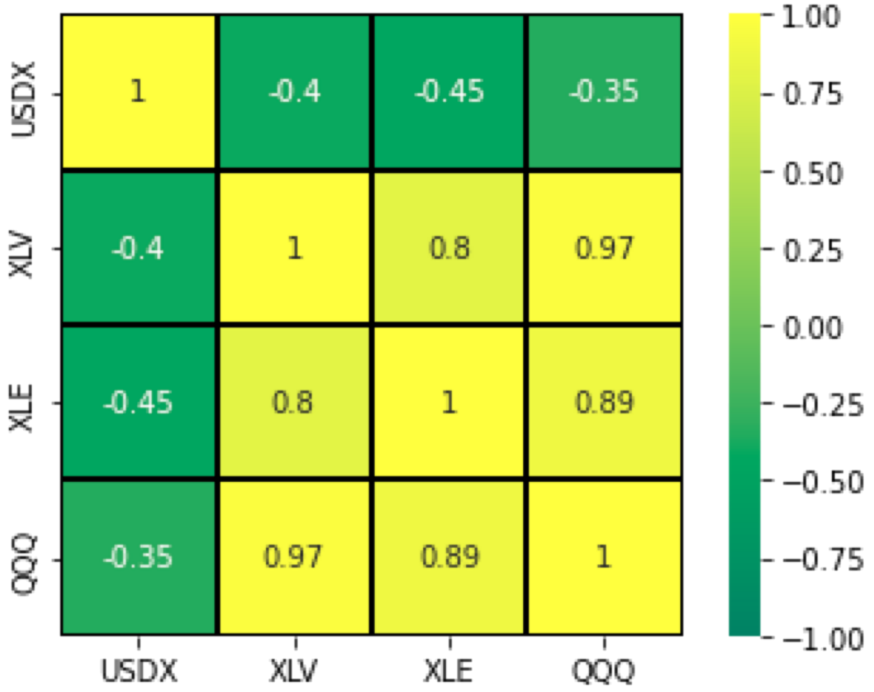
Risk Analysis



Exploratory Data Analysis

Correlation Analysis





2 Different Methods of Analysis

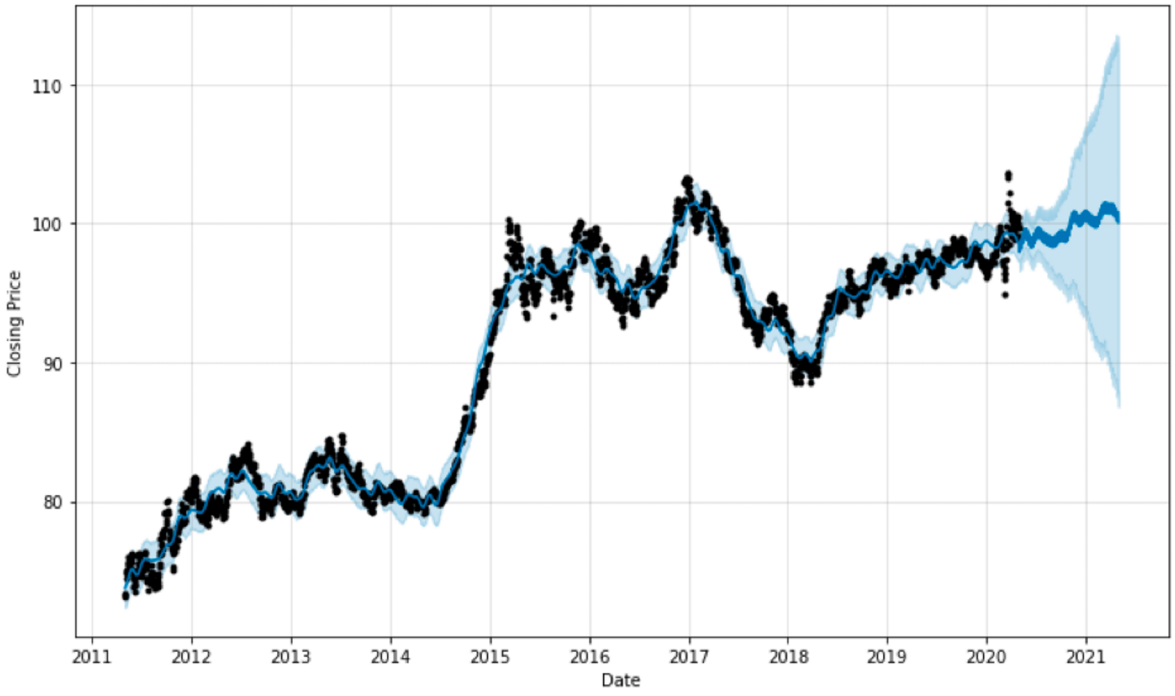
- Prophet Algorithm

- Time Series Analysis - ARIMA model

Forecasting

USDX

- Prophet Algorithm
- Forecasting:100
- Actual price: 91.27
- Error: 9.57%



Part 2: Time Series Analysis

Why we choose Time Series Analysis?

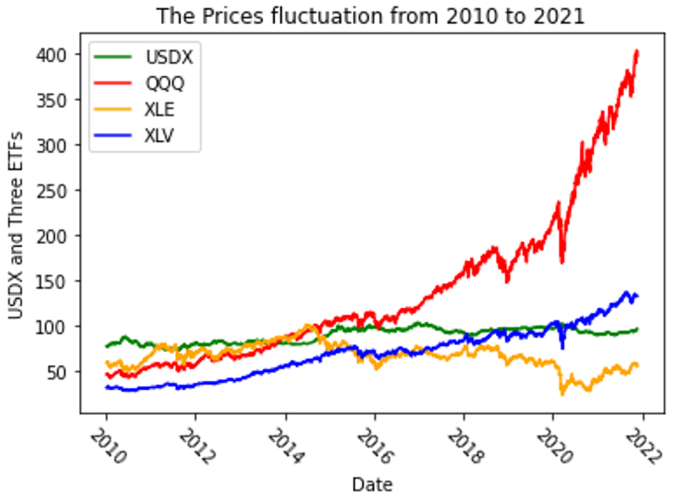
- 2 variables:

  -time

  -the variable we want to forecast

For our data:

- Trend
- Equal interval of time (i.e. daily, weekly or monthly)



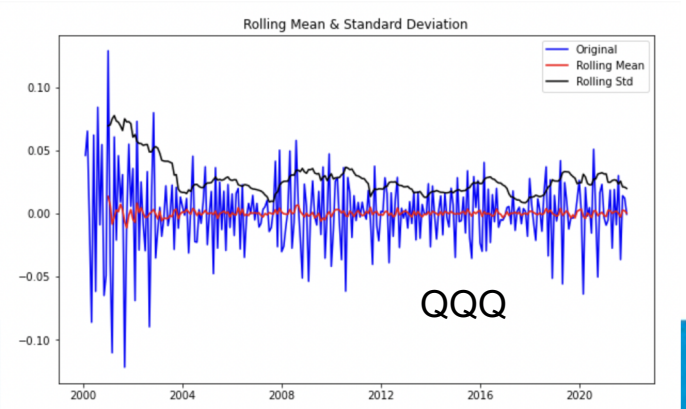
Steps we will take:

Stationary!

1. Remove trend and make time series stationary.
2. Utilize stationarized series with ARIMA model.
3. Transform the data back to the original one.

How do we test for stationary!

ADCF Test - Augmented Dickey–Fuller test

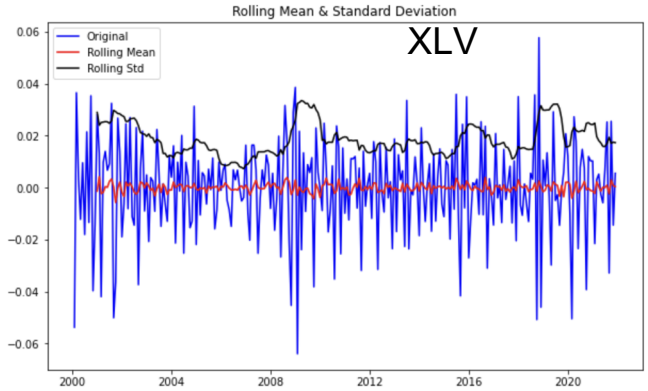



```
Results of Dickey Fuller Test:
Test Statistic                -7.323997e+00
p-value                        1.173998e-10
#Lags Used                     1.400000e+01
Number of Observations Used    2.480000e+02
Critical Value (1%)           -3.456996e+00
Critical Value (5%)           -2.873266e+00
Critical Value (10%)          -2.573019e+00
dtype: float64
```

Null hypothesis: Time Series is not stationary.

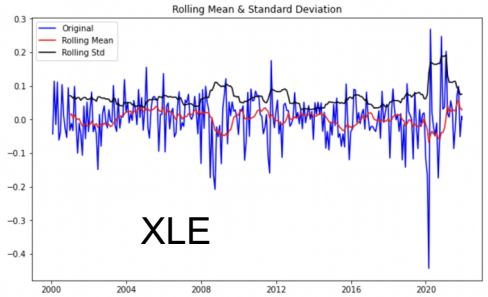
Test Statistic < Critical Value : reject null hypothesis!
P-value is so small: reject null hypothesis!

Another good example of stationary: XLV

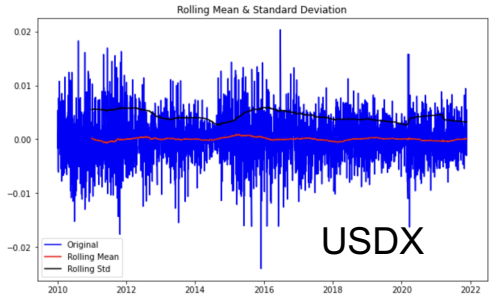



```
Results of Dickey Fuller Test:
Test Statistic                -7.999614e+00
p-value                        2.353154e-12
#Lags Used                     1.400000e+01
Number of Observations Used    2.480000e+02
Critical Value (1%)           -3.456996e+00
Critical Value (5%)           -2.873266e+00
Critical Value (10%)          -2.573019e+00
dtype: float64
```

Two bad examples of stationary: XLE and USDX




```
Results of Dickey Fuller Test:
Test Statistic                -1.668329e+01
p-value                        1.509967e-29
#Lags Used                     0.000000e+00
Number of Observations Used    2.630000e+02
Critical Value (1%)           -3.455461e+00
Critical Value (5%)           -2.872593e+00
Critical Value (10%)          -2.572660e+00
dtype: float64
```




```
Results of Dickey Fuller Test:
Test Statistic                -1.625276e+01
p-value                        3.631480e-29
#Lags Used                     1.000000e+01
Number of Observations Used    2.961000e+03
Critical Value (1%)           -3.432560e+00
Critical Value (5%)           -2.862517e+00
Critical Value (10%)          -2.567290e+00
dtype: float64
```

Another reason for this failure: It is daily data.

Find (p,d,q) for our ARIMA model

ARIMA forecasting equation

- Let Y denote the *original* series
- Let y denote the *differenced* (stationarized) series

No difference (d=0): $y_t = Y_t$

First difference (d=1): $y_t = Y_t - Y_{t-1}$

Second difference (d=2): $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

$= Y_t - 2Y_{t-1} + Y_{t-2}$

Forecasting equation for *y*

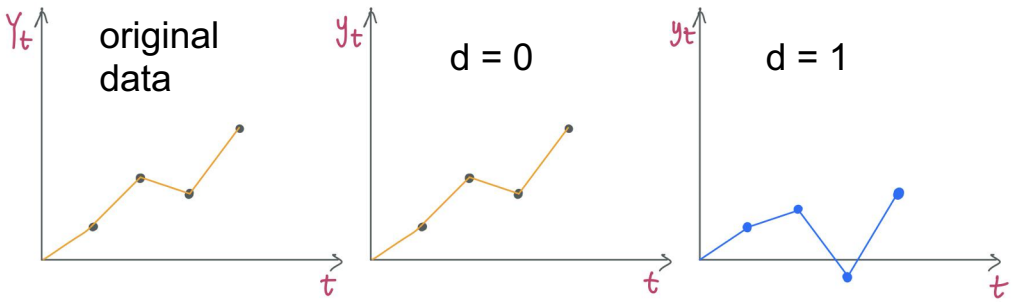
constant — AR terms (lagged values of y)

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} \ldots - \theta_q e_{t-q}$$

MA terms (lagged errors)

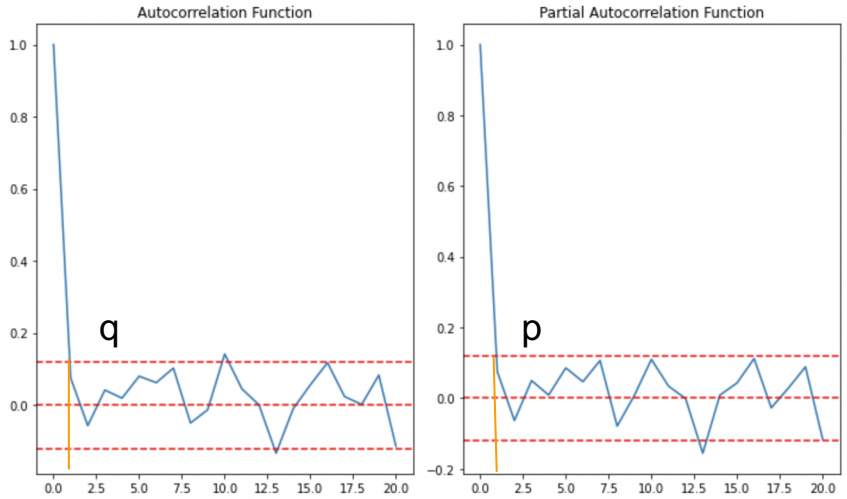
By convention, the AR terms are + and the MA terms are –

Not as bad as it looks! Usually $p+q \leq 2$ and either p=0 or q=0 (pure AR or pure MA model)

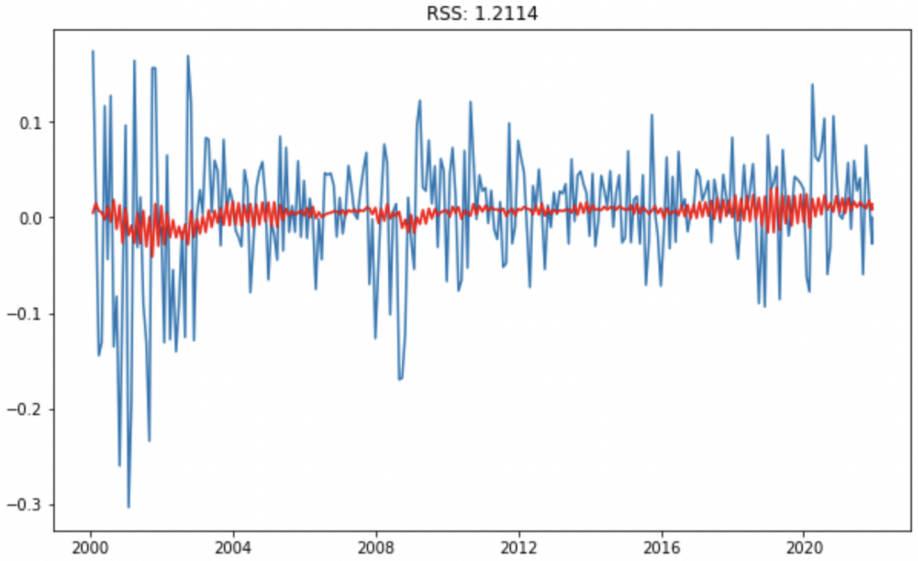


Auto-Regressive Integrated Moving Average

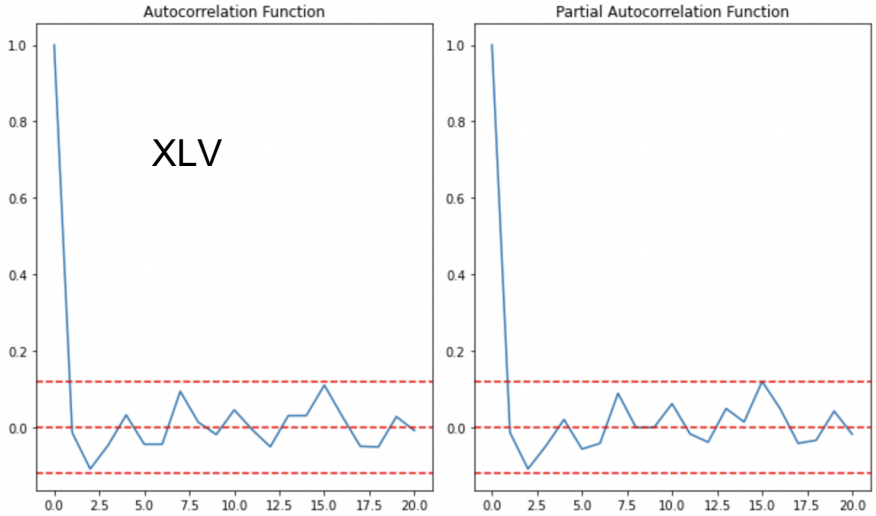
Interpreted by Robert Nau (2014)
https://people.duke.edu/~rnau/Slides_on_ARIMA_models--Robert_Nau.pdf



ACF and PACF for QQQ

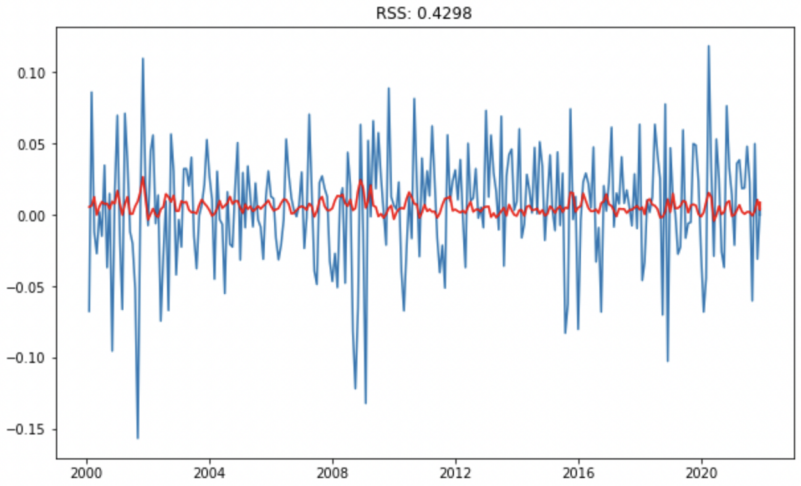



```
# AR+I+MA = ARIMA model
model = ARIMA(indexedQQQ_monthly_logScale, order=(2,1,2))
results_ARIMA = model.fit(disp=-1)
```

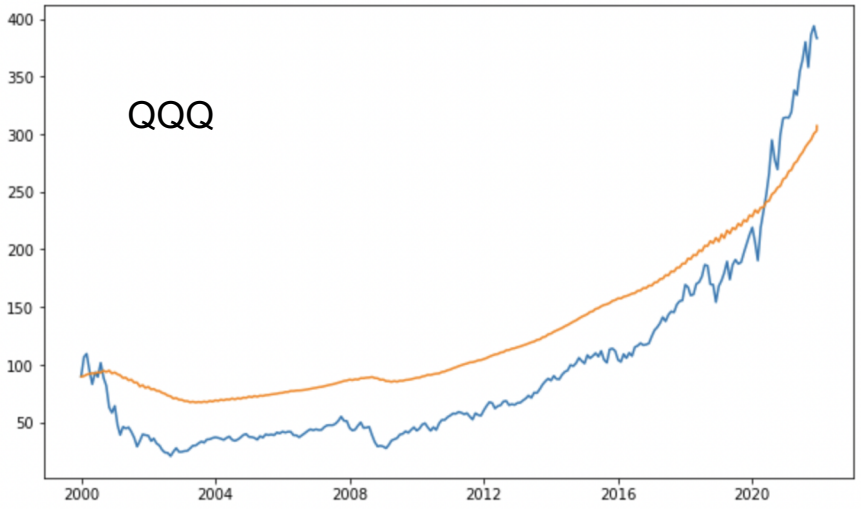


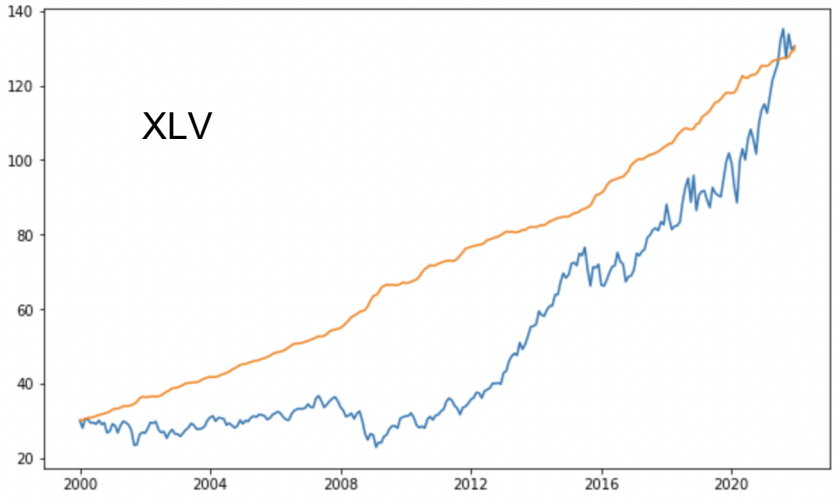
ACF for q and PACF for p

ARIMA model fit for XLV

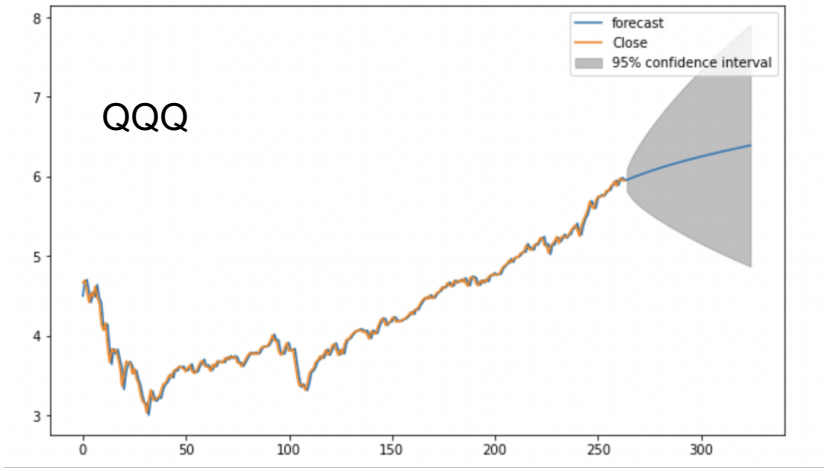


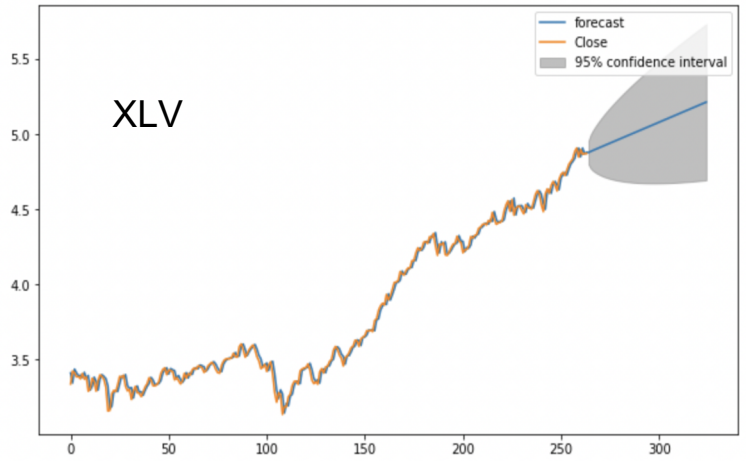
ARIMA model prediction





ARIMA model forecasting





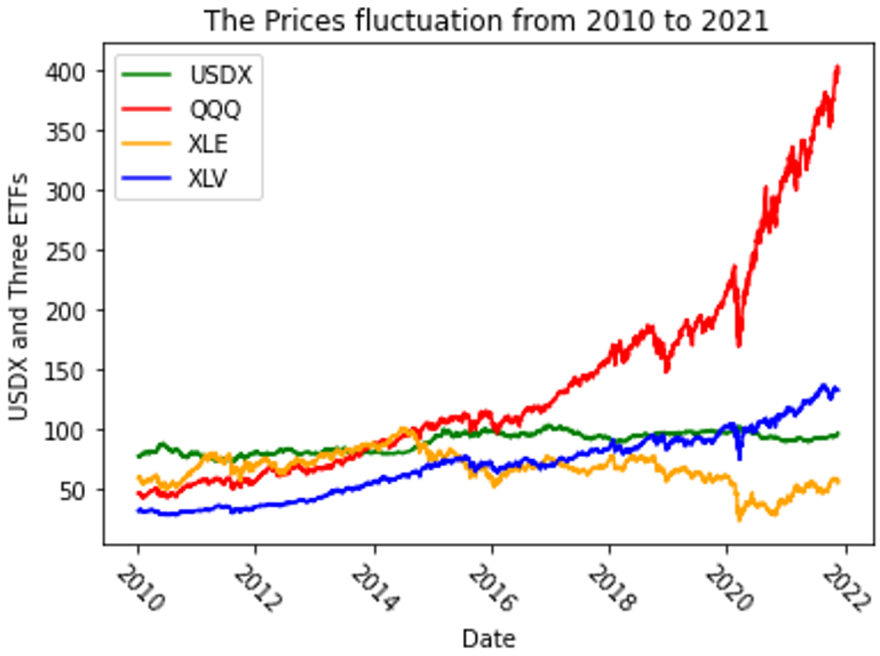
CONCLUSION

- USDX -- Good Performance after Covid-19 Pandemic
- QQQ, XLE, XLV -- Strong Correlation -- Against USDX
- For Risk Averse Investor: USDX + XLV
- For Risk Seeking Investor: USDX + QQQ

- USDX cannot be Time Series and cannot fit ARIMA model
  (no trend)

- QQQ and XLV fit ARIMA very well

- XLE cannot fit Time Series and cannot fit ARIMA model
  (upward and downward, no sign of repetition.



Thank You!