GROUP WORK PROJECT # 2 **MScFE 622: Stochastic Modeling**
GROUP NUMBER: 3573

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Kwan Wei Ma	Hong Kong	kwanweima@gmail.com	
Alper Ülkü	Turkey	alperulku1970@gmail.com	

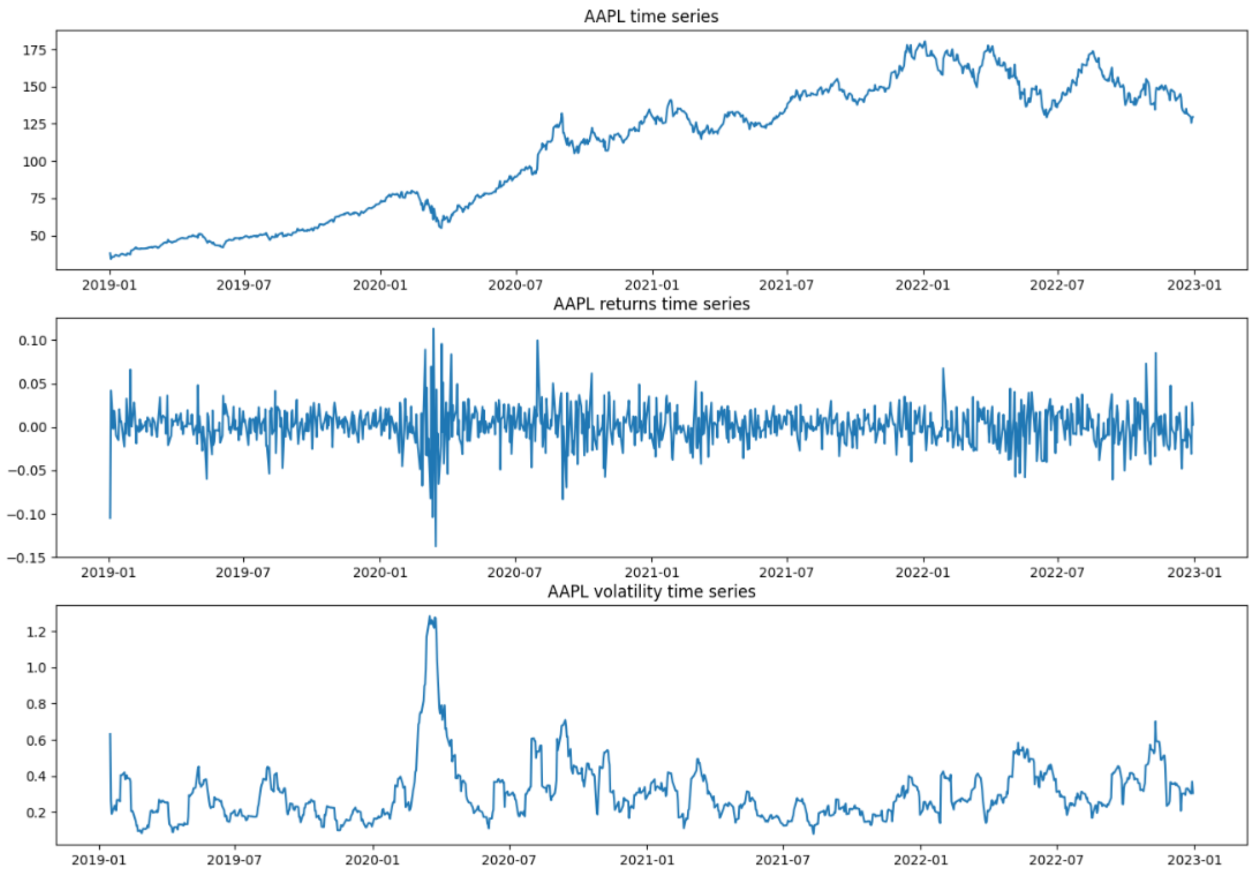
Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	Kwan Wei Ma
Team member 2	Alper Ülkü
Team member 3	

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
Team member 3 was withdrawn by the system.

Step 1: Collecting Financial Time Series

We gathered data for four financial time series: BTC-USD, AAPL, Natural Gas Futures (NG=F), and VIX. All time series cover the period from January 1, 2019, to December 31, 2022 (we prolonged the term to the end of the year and covered a good 4 years). Below is a list of our data. The assets that were chosen are all plotted here to show the different regimes.

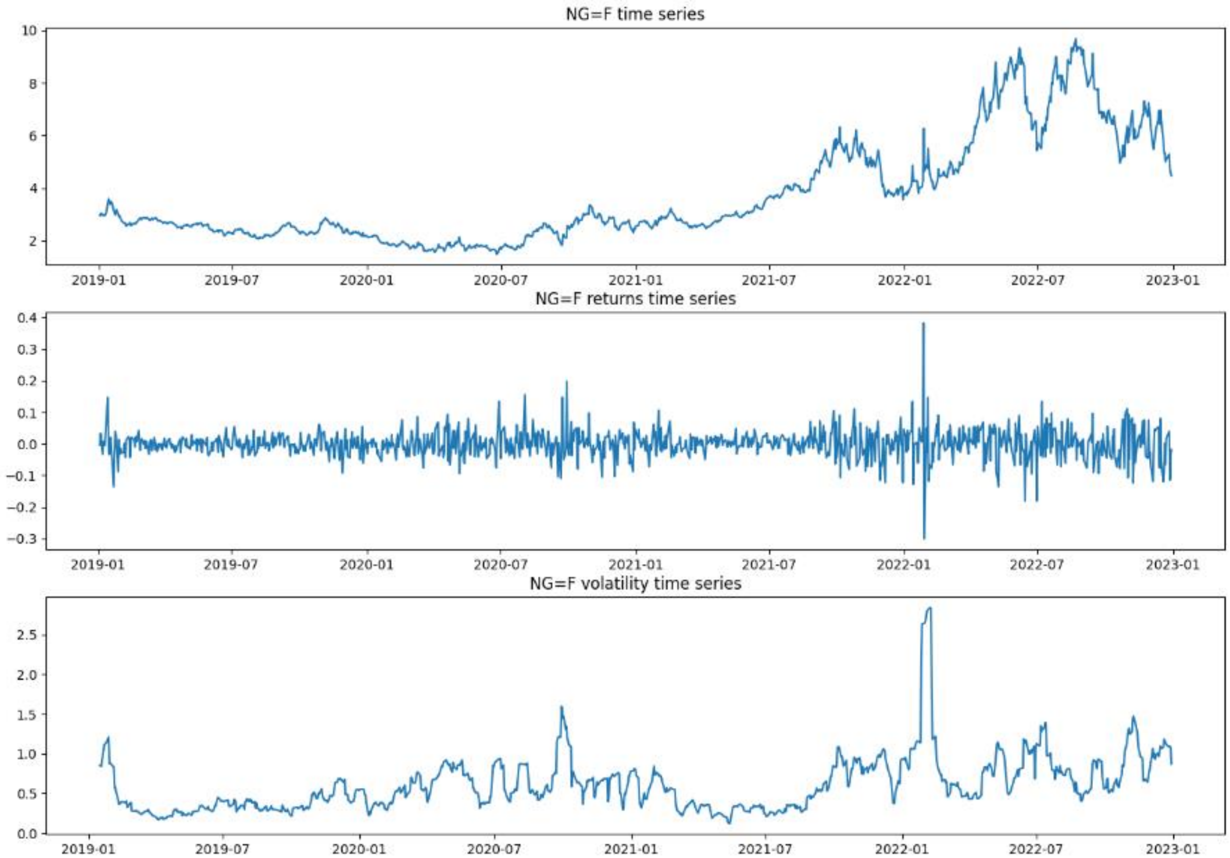
Apple:



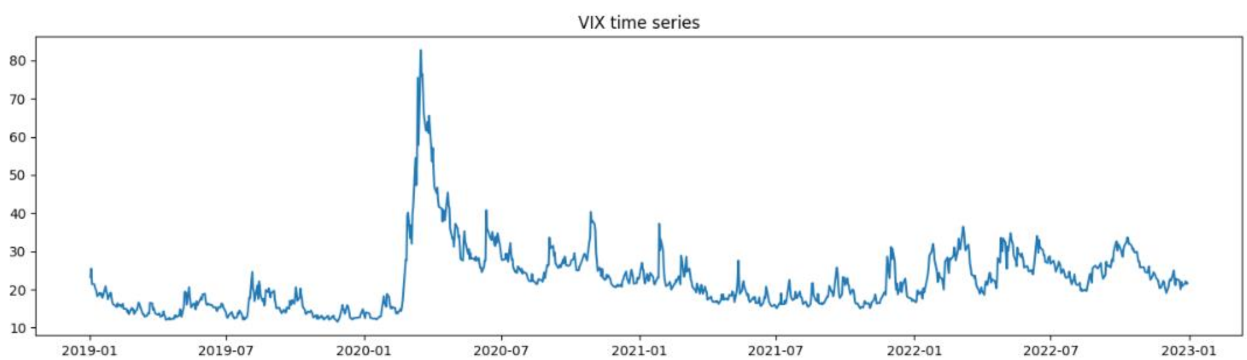
Bitcoin:


BTC-USD time series
70000
60000
50000
40000
30000
20000
10000
2019-01
2019-07
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
BTC-USD returns time series
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
BTC-USD volatility time series
2.5
2.0
1.5
1.0
0.5
0.0

Natural Gas Futures:



VIX:



Step 2a. Data selection

We can see that there have been multiple regime transitions from each of the assets. The most drastic regime changes, however, took place in February 2022 and March 2020. During the first week of March, while the VIX index was starting to break its yearly high, around 80, the AAPL volatility swung up to around %120, BTC was to %250 on March 2020 and NG=F swung over %270 in Feb 2022.

Among these 4 data, we think all can be used for volatility modelling and estimation but we selected VIX as it is a common universal data set for disseminating volatility and VIX seemed quite capable of showing the significant volatility change within post-COVID19 markets.

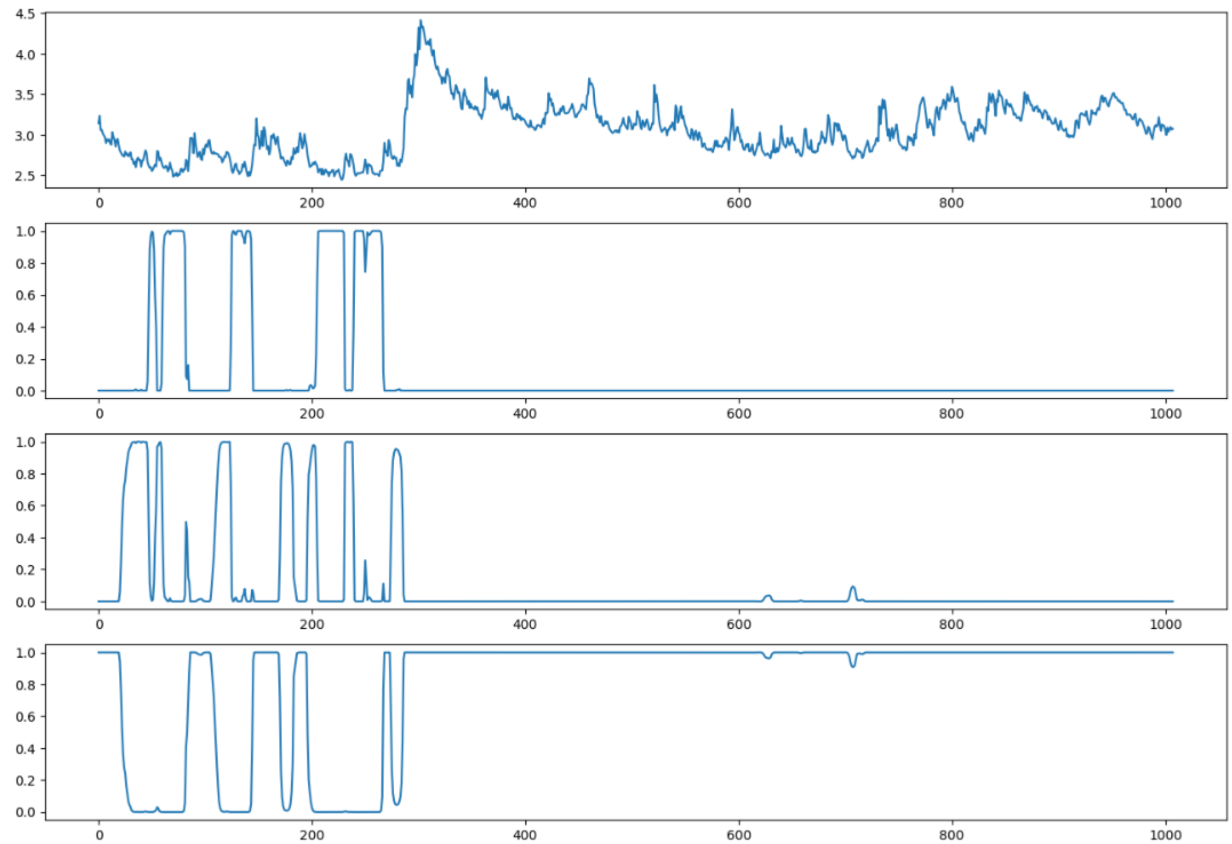
In this report, we will estimate a Markov-regime switching model for the VIX index, and compare the performance of different models under different assumptions using AIC, and finally estimate a Markov-regime switching model assuming the VIX index time series arises from the AR process.

STEP 2b.

Case 1: N=3, different mu and sigmas.

Initial States:

```
mu_hat0 = [2.5 3.  3.5],
sigma_hat0 = [0.3 0.4 0.5],
P_hat0 =
[[0.85  0.075 0.075]
 [0.05  0.9   0.05 ]
 [0.15  0.15  0.7  ]],

pi_hat0 = [0.3333 0.5    0.1667]
```



Markov-regime switching model output for VIX with n = 3, different mus, and sigmas

When we run the code with log of VIX data taken in 67th iteration calculation converges with following parameters:

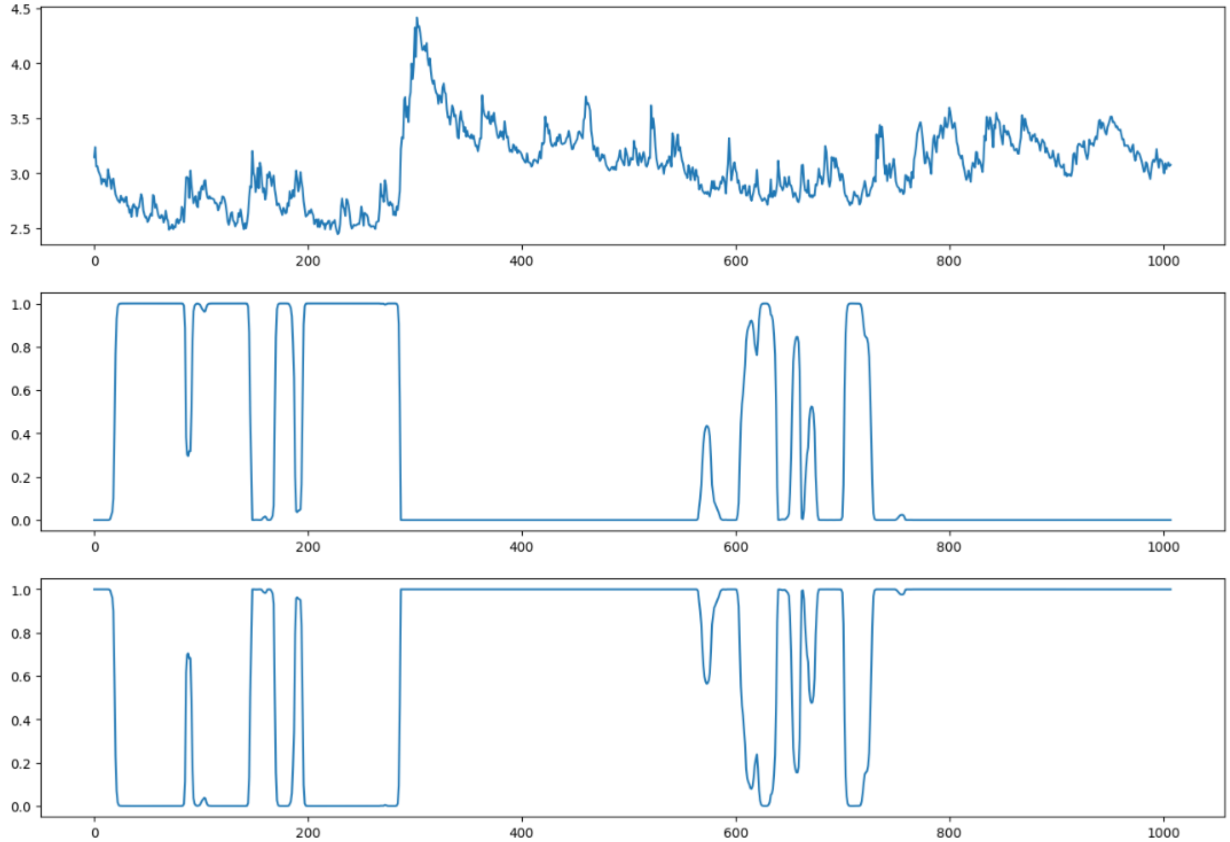
```
ite=67, max(diff) = 0.0072
mu_hat1 = [2.5574 2.7017 3.1517],
sigma_hat1 = [0.0483 0.0556 0.2842],
P_hat1 =
[[0.944  0.0319 0.0242]
 [0.0648 0.9032 0.032 ]
 [0.     0.0064 0.9936]],

pi_hat1 = [0. 0. 1.]
```

COVID volatility significantly detected but post COVID dates volatility changes went almost undetected.

Case 2: n=2, start with different mu, same sigma.

Initial States:

```
mu_hat0 = [2, 4]
sigma_hat0 = [0.1, 0.1]
```



When we run the code with log of VIX data taken in 29th iteration calculation converges with following parameters:

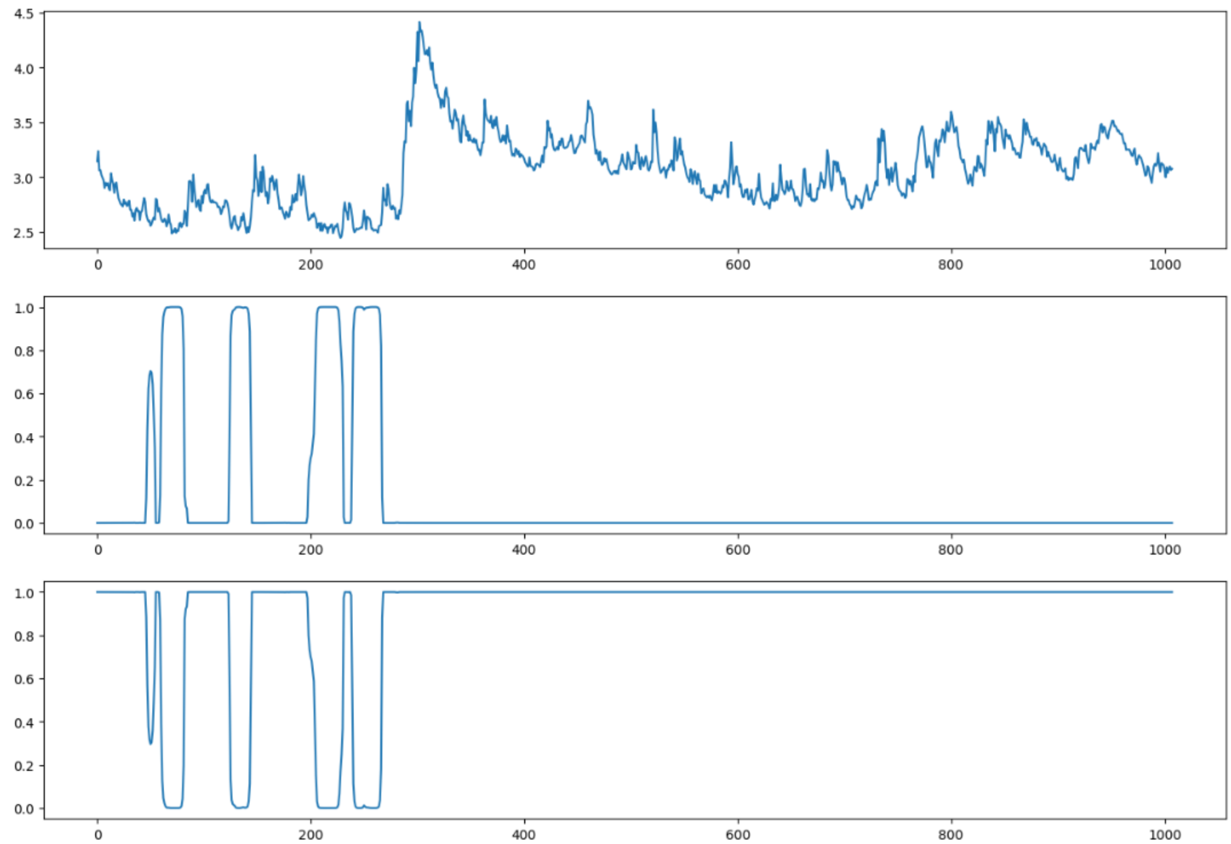
```
ite=29, max(diff) = 0.1812
mu_hat0 = [2.6994 3.2142],
sigma_hat0 = [0.1228 0.2633],
P_hat0 =
[[0.9745 0.0255]
 [0.0116 0.9884]],

pi_hat0 = [0. 1.]
```

We had better stopped the iteration at 29th as we will be able to detect regime change after post COVID 19 dates. If we did not, regime change after COVID 19 dates may not be detected. In this case COVID volatility significantly detected and also post COVID dates volatility changes were detected as well.

Case 2: n=2, start with different mu, different sigma.

Initial States:

```
mu_hat0 = [2. 4.],
sigma_hat0 = [0.1 0.2],
P_hat0 =
[[0.85 0.15]
 [0.1  0.9 ]],

pi_hat0 = [0.4 0.6]
```



When we run the code with log of VIX data taken in 23rd iteration calculation converges with following parameters:

```
ite=23, max(diff) = 0.00902

mu_hat1 = [2.5604 3.1085],
sigma_hat1 = [0.0502 0.3019],
P_hat1 =
[[0.9527 0.0473]
 [0.0052 0.9948]],
pi_hat1 = [0. 1.]
```

COVID volatility significantly detected but post COVID dates volatility changes went almost undetected.

Step 3: Performance Comparison

We will use the VIX model as an illustration in this section. AIC will be used to evaluate each model's performance. We keep in mind that AIC = -2(Log-likelihood) + 2(number of model parameters).

Different mus

First, for n = 2, we determine the log-likelihood and AIC of the models with the same mus and different mus. The following outcomes are attained:
- The VIX model's log-likelihood with n=2 and the identical mus is 1430.29, and its AIC is -2836.58.
- The VIX model's log-likelihood with n=2 and different mus is 1443.63, and its AIC is -2863.27.

From the result, the model under assumption of different mu values has a lower AIC, showing a better performance.

Different sigmas

Then, for n=2 and n=3, respectively, we compute the log-likelihood and AIC of the models with various sigmas. The following outcomes are attained:
- The VIX model's log-likelihood with n=2 and the same sigmas is 1430.29, and its AIC is -2836.58.
- The VIX model's log-likelihood with n=2 and different sigmas is 1430.16, while the AIC is -2836.32.

As a result, the model with various sigma values has a lower AIC, demonstrating greater performance.

AIC comparisons:

With L = 1430.290039, AIC score is -2836.58, n=2
With L = 1443.633996, AIC score is -2863.27, n=2
With L = 1430.159694, AIC score is -2836.32, n=2
With L = 1443.642587, AIC score is -2863.29, n=2
With L = 1493.535963, AIC score is -2963.07, n=3

Lowest AIC score was achieved when N = 3 and different mus and sigmas.

Conclusion

According to the findings, models with distinct mus are preferable since their AIC values are lower. Overall, the VIX model with n = 3 and various mus is the best model, followed by the model with n = 2. The poorest VIX models are those with varying sigmas since they all have the same AIC.

We also observe that the AIC differences between the VIX models were close to ~ 30, which indicating evidence for less difference in Markov model in each case. However, for the models with different mus and sigma and n=3, the AIC difference is greater than ~ 100, which indicates that the updated model is less similar to the models with n=2.

Step 4: Estimate the model from the AR process

In this section, we will estimate the model under the assumption that the time series results from an autoregressive process, with the variance of the perturbation term and the autoregressive coefficient changing with the state.

First we should test whether our data is stationary or not. We need an ADF test for this.

In our code we used:

```
from statsmodels.tsa.stattools import adfuller
# ADF Test
X = YDatapd.dropna().values
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

so that this yielded:

```
ADF Statistic: -3.857285
p-value: 0.002373
Critical Values:
      1%: -3.437
      5%: -2.864
      10%: -2.568
```

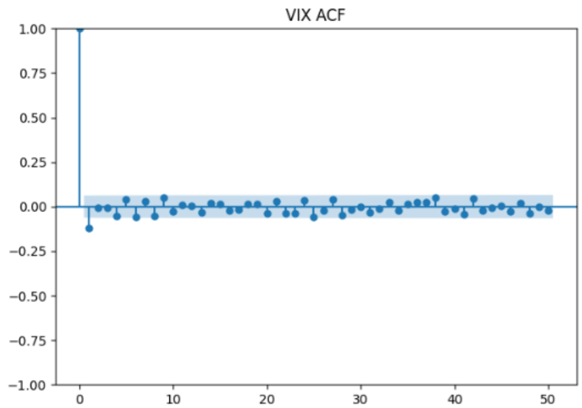
As we observe that p-value is less than 0.05 we reject the null hypothesis on non-stationarity and detect that our data is stationary with %99 fidelity, as we observe that test statistic = -3.857 is less than -3.437.

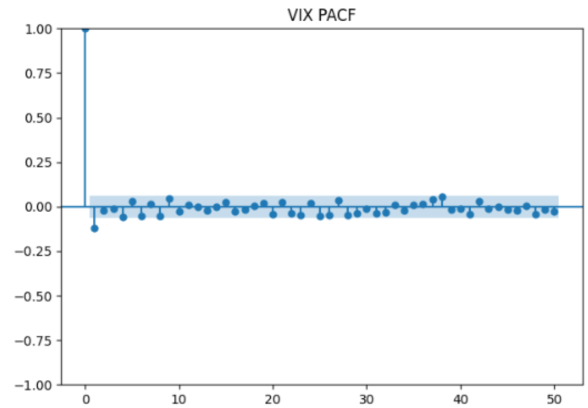
Thus we safely continued our analysis with ACF and PACF tests:

We coded:

```
#Plot ACF and PACF graph
# plot ACF and PACF
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 5))
sm.graphics.tsa.plot_acf(np.log(YDatapd).diff().dropna(), title="VIX ACF",
lags=50, ax=ax1)
sm.graphics.tsa.plot_pacf(np.log(YDatapd).diff().dropna(), title="VIX
PACF", lags=50, ax=ax2)
plt.show()
```

yielding:





As seen above in ACF and PACF graphics, VIX data seems to be not so dependent on past values, the most significant dependence is to its previous value, just one sample before. This happily matched with the Markov property.

From the above graph, it can be seen that the VIX time series follows a first-order autoregressive process. Therefore, we are going to continue our analysis assuming it is an AR(1) process.

We run the ARIMA(1,1,0) model to find the autoregressive coefficient.

We coded and resulted as:

```
# ARIMA Model for VIX
mod_can_a = ARIMA(
    np.log(YDatapd), order=(1, 1, 0), trend="n"
).fit()
print(mod_can_a.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1008
Model:                 ARIMA(1, 1, 0)   Log Likelihood                1114.385
Date:                Tue, 01 Aug 2023   AIC                          -2224.769
Time:                        13:13:46   BIC                          -2214.940
Sample:                             0   HQIC                         -2221.035
                               - 1008
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1181      0.023     -5.066      0.000      -0.164      -0.072
sigma2         0.0064      0.000     39.955      0.000       0.006       0.007
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):              1055.82
Prob(Q):                              0.93   Prob(JB):                         0.00
Heteroskedasticity (H):               0.73   Skew:                             1.25
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.35
===================================================================================
```

From the results, as seen above SARIMAX window, the autoregressive coefficient would be -0.1181 and sigma would be 0.0064.

Next we would make use of the output to calculate the transition probabilities using the Tauchen method. Tauchen method is a numerical technique used to discretize a continuous-valued stochastic process into a finite set of states.

The transition probabilities are given by below calculation:

$$p_{ij} = \Phi\left(\frac{m_j - \rho z_i}{\sigma_\varepsilon}\right) - \Phi\left(\frac{m_{j-1} - \rho z_i}{\sigma_\varepsilon}\right) \text{ for } j = 2, \ldots, N-1$$

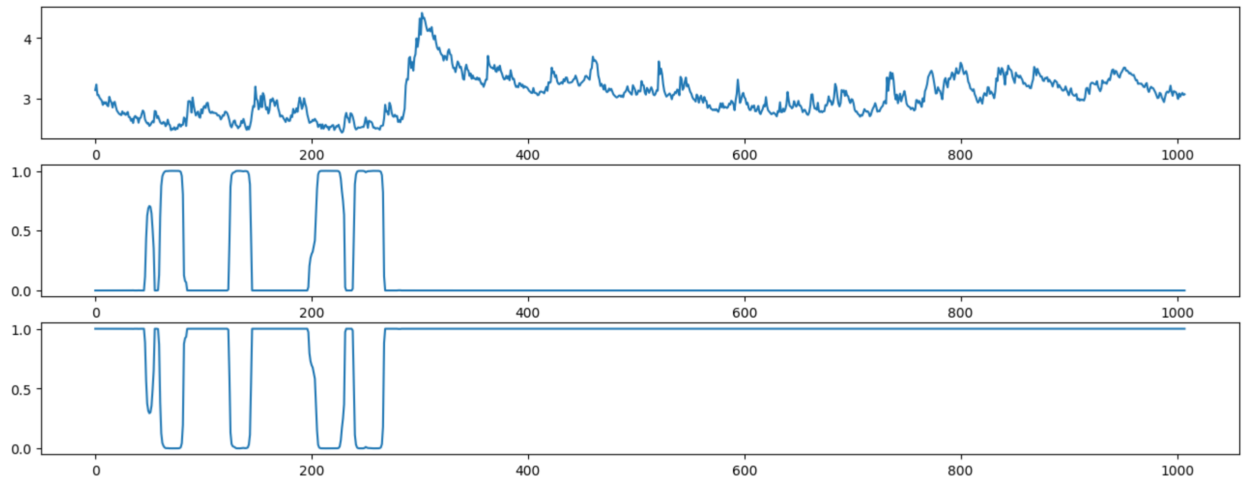
$$p_{i1} = \Phi\left(\frac{m_1 - \rho z_i}{\sigma_\varepsilon}\right)$$

$$p_{iN} = 1 - \Phi\left(\frac{m_{N-1} - \rho z_i}{\sigma_\varepsilon}\right)$$

```
# Fix the parameters using arima model output
RHO = -0.1181
SIGMA = 0.0064
N_GRID = 2
LAMBDA = 2.0 # Used in Tauchen method

P_tauchen, zgrid_tauchen = tauchen_method(RHO, SIGMA, LAMBDA, N_GRID)
print(P_tauchen)
```

We set lambda = 2 and run the code for the Tauchen method to calculate the transition probabilities.

Finally, we estimate the Markov-regime switching model using the calculated transition probabilities.



When we run the code with log of VIX data taken in 10th iteration calculation converges with following parameters:

```
ite=10, max(diff) = 0.0094
mu_hat1 = [2.5604 3.1085],
sigma_hat1 = [0.0502 0.3019],
```

```
P_hat1 =
[[0.9527 0.0473]
 [0.0052 0.9948]],
pi_hat1 = [0. 1.]
```

COVID volatility significantly detected but post COVID dated volatility changes went almost undetected.

References

1. Stochastic Modeling Course Notes. WorldQuant University, 2023.
2. Financial Econometrics Course Notes. WorldQuant University, 2023.
3. “Akaike’s Information Criterion: Definition, Formulas.” Statisticshowto, statisticshowto.com/akaikes-information-criterion.