GROUP WORK PROJECT # GWP2M5 MScFE 24/06 622 Stochastic Modeling

GROUP NUMBER: 6524

FULL LEGAL NAME	**LOCATION (COUNTRY)**	**EMAIL ADDRESS**	**MARK X FOR ANY NON-CONTRIBUTING MEMBER**
Shong Xian Lee	Singapore	shongxian5601@gmail.com	
Emmanuel Hansingo	Zambia	emmanuel.hansingo@gmail.com	
Ketty Muwowo	Zambia	kettymuwowo@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	Shong Xian Lee
Team member 2	Emmanuel Hansingo
Team member 3	Ketty Muwowo

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.

Contents

Step 1

Team member A (Shong Xian) is preparing the financial time series for the equity category which is IBM. The time frame was scheduled from 1st January 2019 to 31st December 2022 that covered the whole process of covid. It will focus on the time series of the close price and identify the possible ways to deal with this fast-changing event.

Team Member B (Emmanuel Hansingo) selected the S&P 500 index as the time series dataset to use to observe regime change for the period between 2019 and 2020. This is because a dataset like S&P 500 can help detect shifts in economic policies, market conditions, and structural changes.

Team Member C (Ketty Muwowo) selected the Bitcoin closing price data for analysis from the cryptocurrencies time series. The chosen time frame was from the beginning of 2019 in the pre COVID times up to 31st December 2022 . The Bitcoin data set captured significant market movements during this period and sharp declines were observed in 2020. There was also a subsequent recovery and some continued volatility throughout 2021 and 2022 as shown from the Bitcoin time series graph in step 2.

Step 2

IBM Data

Hidden Markov Models

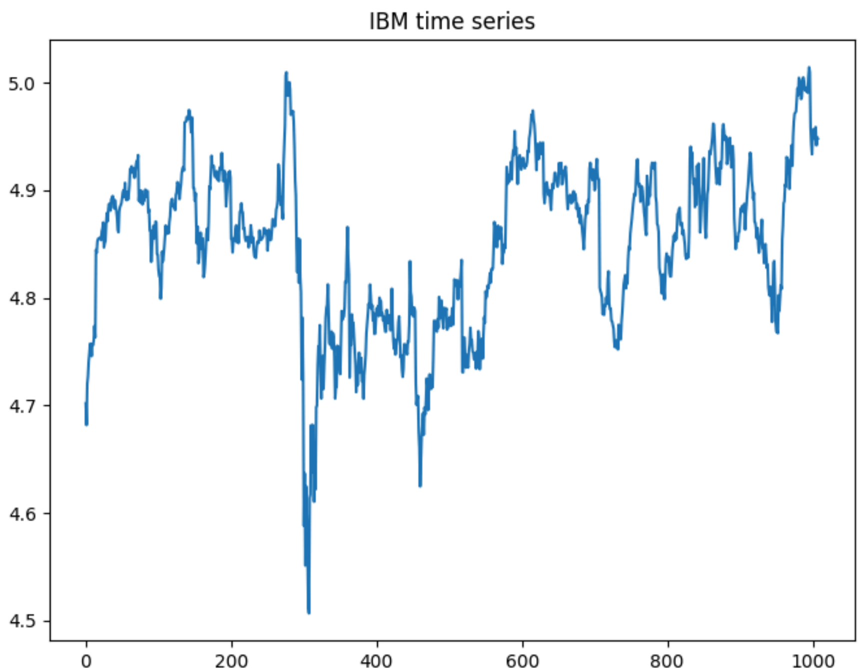
It is part of the Markov process that is split into two components which are observable and unobservable components Unobservable component refers to hidden Markov model. It describes the stochastic system is determine by a set of components which are (WorldQuant University, 2024):

a. A set of N states S
b. Transition probability matrix P
c. A sequence of T with possibly vector-valued
d. A sequence of observation marginal likelihood for each i
e. An initial probability distribution of pi.

The assumption of the hidden Markov process is independent of past observation.

$$\mathcal{Y}_{t-1}, \text{ i.e., } \mathbb{P}\{s_t = j | s_t = i, \mathcal{Y}_{t-1}\} = \mathbb{P}\{s_t = j | s_t = i\} = p_{ij}.$$

Regime Change of IBM's Close price (Time series)



The chart showing above describes the frequency of price changes. Y-axis indicates the range of volatile from 4.5 to 5 over the period during covid-19. X-axis indicates the number of days that ranges from 01 January 2019 to 31 December 2022. A steep decrease appeared around 300 days. This decrease indicates the outbreak of covid-19 which is close to the year end of 2019.

Parameter assumption

```
Markov-regime switching model (state 2)

# Initialize parameters
T = len(YData)
N = 2

# SET INITIAL GUESSES

mu_hat0 = [2, 4] * np.ones((N))
sigma_hat0 = [0.1, 0.1] * np.ones((N))
P_hat0 = np.zeros((N, N))

P_hat0[0, 0] = 0.75
P_hat0[0, 1] = 1 - P_hat0[0, 0]
P_hat0[1, 1] = 0.7
P_hat0[1, 0] = 1 - P_hat0[1, 1]

pi_hat0 = [0.5, 0.5] * np.ones((N))
for t in range(1, 100):
    pi_hat0 = np.dot(P_hat0.T, pi_hat0)
```

It assumes the number of states is equal to 2. The initial guess of the expected values (means) are 2 and 4 were set to each state and times with number of states. The variance of the time series in each state is set at 0.1 in each state. The state transition probabilities were based on the number states, 75% probability is set in state 0.

25% of probability is set from state 0 to state 1. 70% of probability transition state in state 1 and 30% probability from state 1 to 0. Since the number of states is 2, the vector of initial state probabilities for the Markov regime-switching model is set to 0.5.

Number states of 3

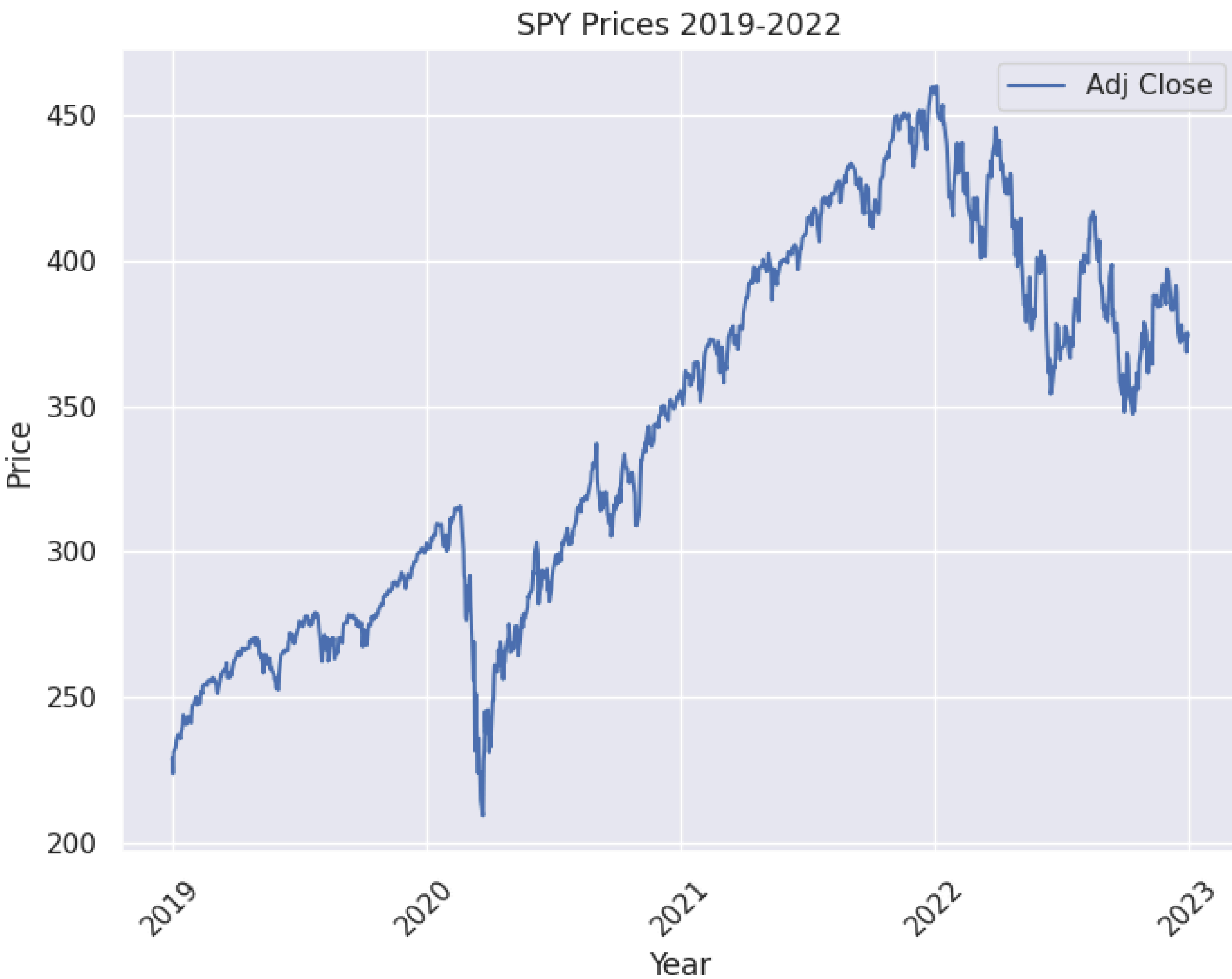
```
# Initialize parameters
T = len(YData)
N = 3

# SET INITIAL GUESSES
mu_hat0 = np.array([2, 4, 6])
sigma_hat0 = np.array([0.1, 0.1, 0.1])
P_hat0 = np.array([[0.75, 0.15, 0.1], [0.2, 0.7, 0.1], [0.1, 0.1, 0.8]])
pi_hat0 = np.array([0.33, 0.33, 0.33])
for t in range(1, 100):
    pi_hat0 = np.dot(P_hat0.T, pi_hat0)
```

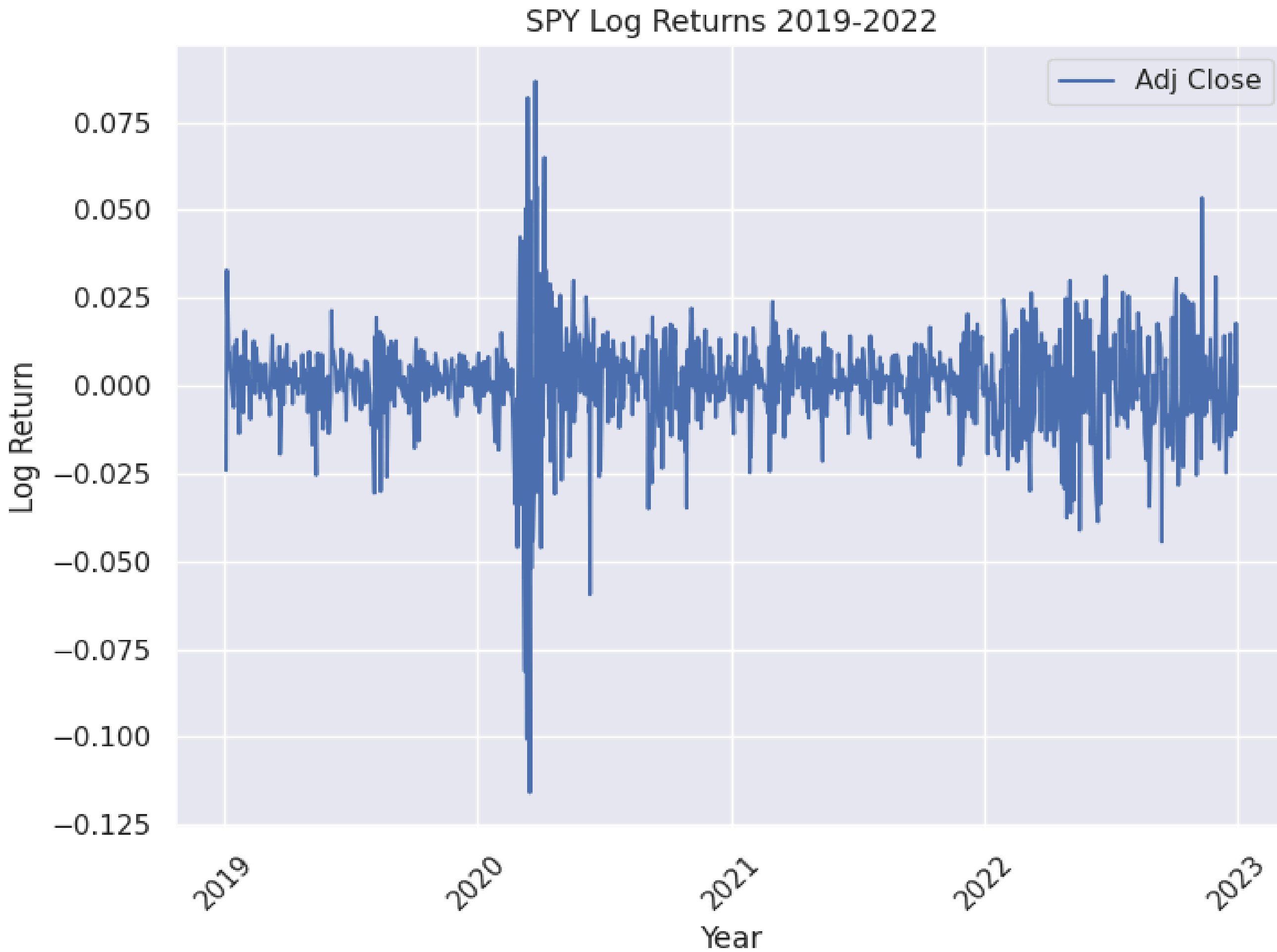
The number states are set to 3. The initial guess of the expected mean values has been set to 2, 4, and 6. Each states' variance of the time series has been set at 0.1 which is similar to the previous state. A 75% probability is staying in state 0 and 15% to state 0 to state 1. The rest of 10% was set to state 1 to state 2.

The probability of the second row was set to 20%, 70%, and 10%. The third row is set to 10%, 10%, and 80%. Since the number of states is 3, the vector of initial state probabilities for the Markov regime-switching model is set to 0.33.

S&P 500

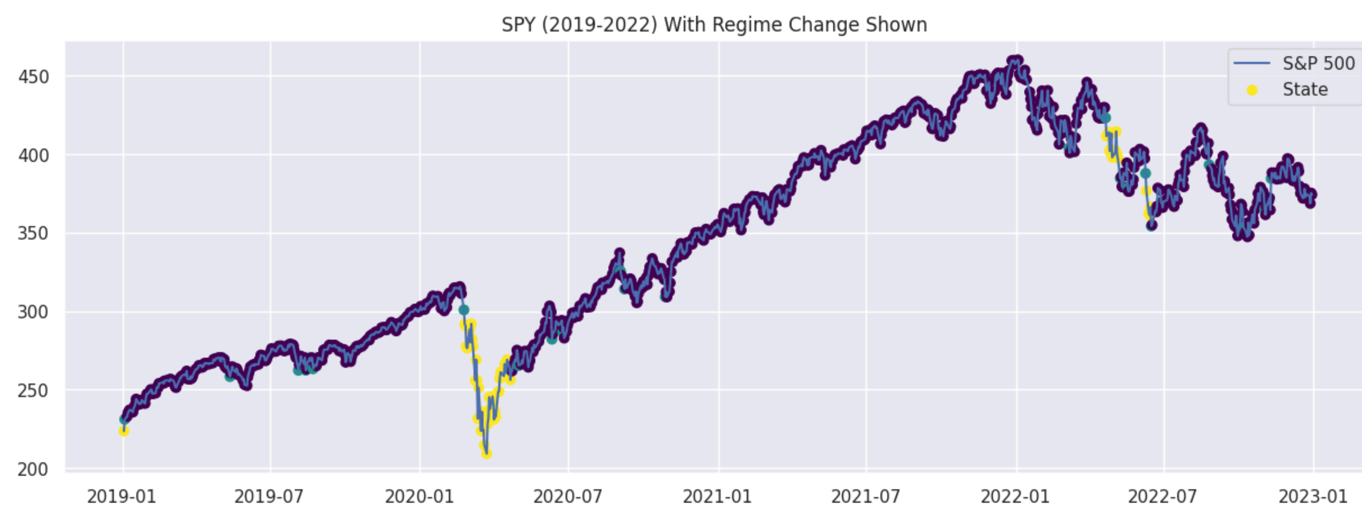


The data shows an abrupt drop in the price in 2020 around the height of the COVID pandemic. This is expected given the high volatility due to the uncertainty in the markets in that period.



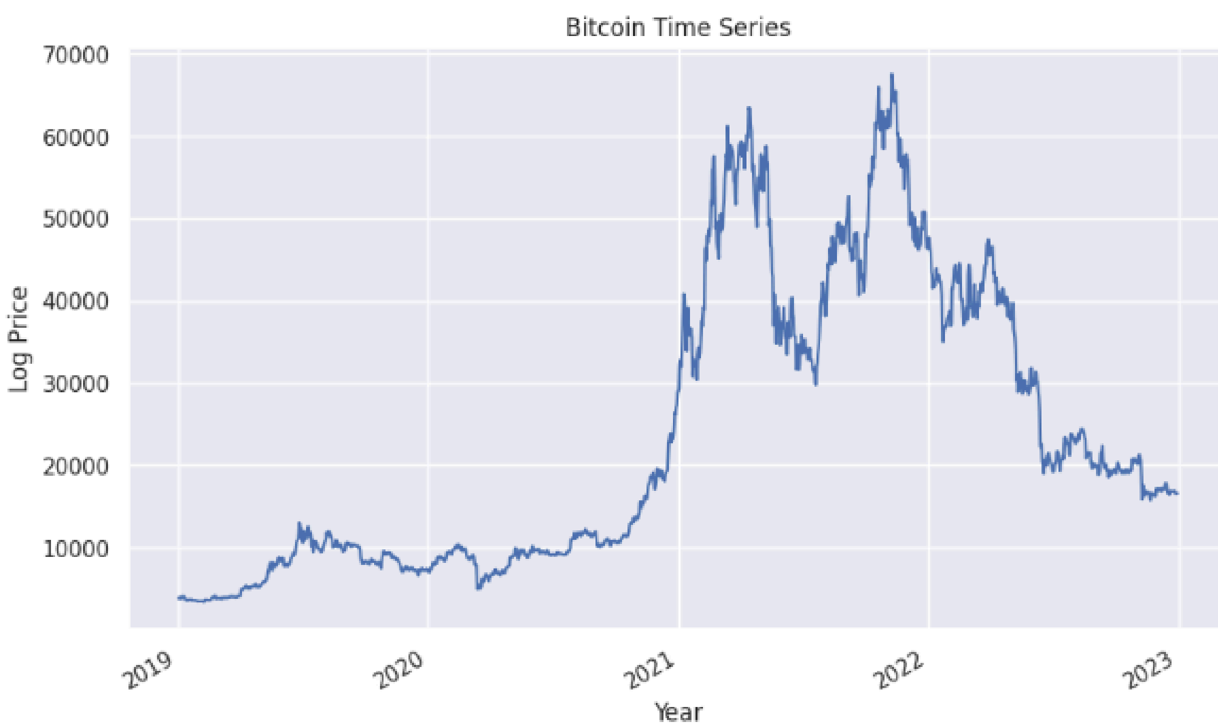
The high volatility is also observed in the returns of the index during the year 2020. This volatility is also observed to a lesser degree between 2022 and 2023

State	Adj Close	Returns
0	347.316012	0.001458
1	326.422350	-0.023163
2	290.775506	-0.005437



Regime changes observed in the price data. Regime changes refer to shifts in the underlying processes governing a system; they are used in economics and finance to describe changes in economic policy, market conditions, or structural shifts. In time series models like the one used here, this can involve changes in the mean, variance, or other parameters over time. These shifts can be abrupt as seen in 2020 or gradual, and accounting for them is crucial for accurate modeling and forecasting.

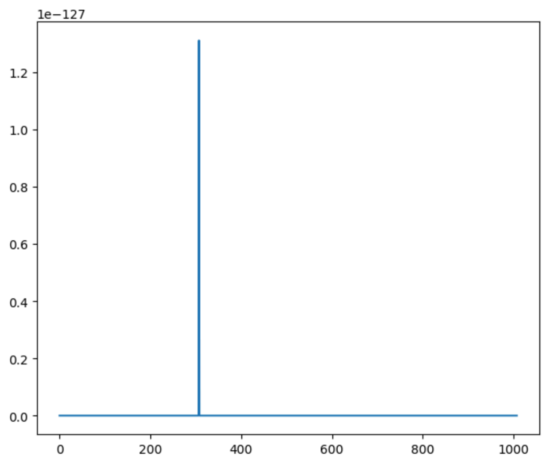
Bitcoin

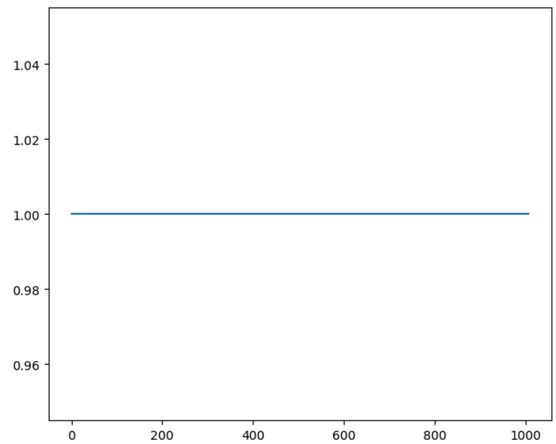


Step 3

Team member A

Number state of 2






```
4 4.0763088344285916e-05
[4.50647954 4.8498628 ] [0.         0.07826506] [[0.00000000e+000 1.00000000e+000]
 [1.30205082e-130 1.00000000e+000]] [0. 1.]
```

The value of the convergence check is 0.0000408 that indicates the maximum difference between existing and new estimates of parameters like number of states, mean, variance, transition probabilities. The smaller the better which describes the expectation-maximization algorithm has almost converged and does not require further iteration. This also means that the estimates become accurate.

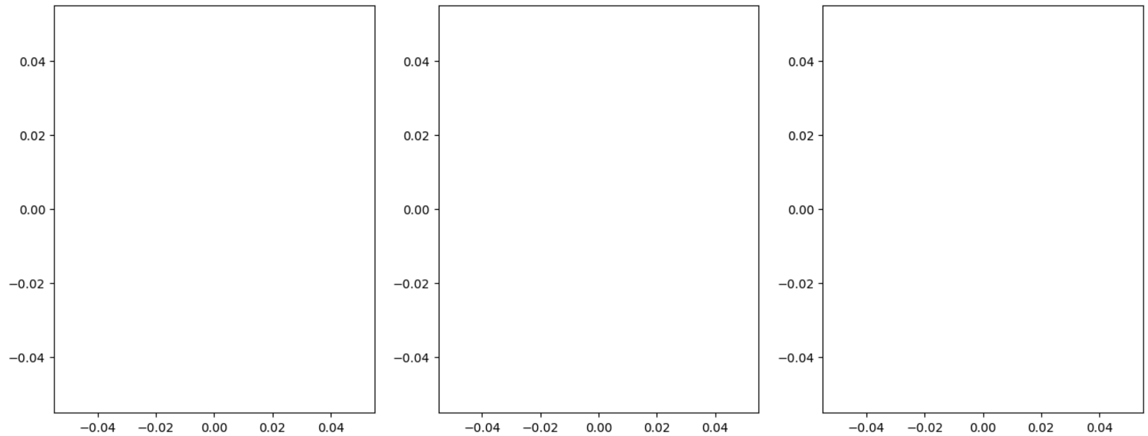
The mean value is described at 4.51 and 4.85 under 2 different states. This indicates both mean values are close enough to each state. The variance of the first state is 0 and the second state is 0.078. The probability of transitioning into the second state is 1 and the probability of staying in the first state is 0. This means the transition process remains in the second state.

Team member B

Constant Mu and Varying Sigma

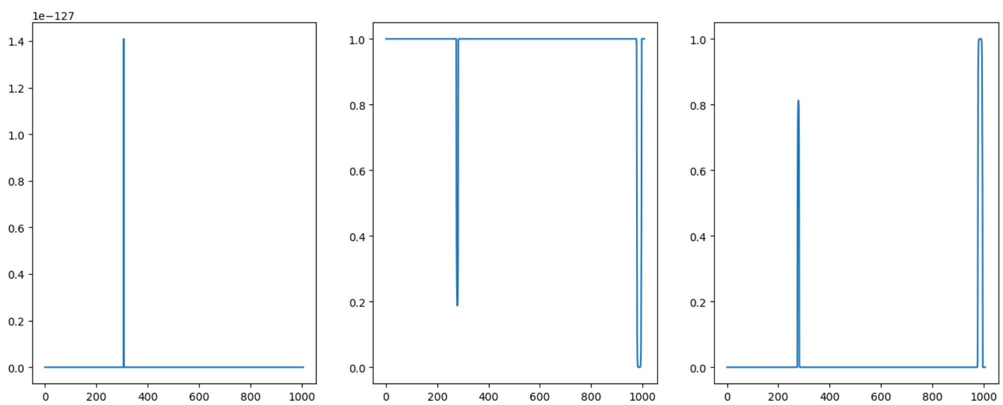
The model used a constant mean of 2 while having a varying variance over two states. The model with varying variance is better suited for time series with evident changes in volatility. It adds complexity due to additional parameters (different variances for each regime). With changing variance, the model can better capture different volatility regimes in the data. COmpared to a model with constant variance, the model may not accurately capture changes in volatility, leading to less accurate regime identification.

Number state of 3




```
5 nan
[nan nan nan] [nan nan nan] [[nan nan nan]
 [nan nan nan]
 [nan nan nan]] [nan nan nan]
```

The iteration step was stopped at 5 and given output of nan value. This indicates numerical instability. Hence, we decided to refer back to iteration at step 4.




```
4 60.56612790453337
[4.50647954 4.84644926 4.99716309] [0.         0.07583833 0.00791264] [[0.00000000e+000 1.00000000e+000 0.00000000e+000]
 [1.43202213e-130 9.98151091e-001 1.84890902e-003]
 [0.00000000e+000 7.97028502e-002 9.20297150e-001]] [0. 1. 0.]
```

Based on the output, the value of convergence check is 60.57 that indicates the maximum difference between existing and new estimates of parameters like number of states, mean, variance, transition probabilities. This higher value indicates the expectation-maximization algorithm has not fully converged and required further adjustment on the parameters.

The mean value is described at 4.51, 4.85, and 5 under 3 different states. This indicates these 3 mean values are close enough to each state. The variance of the first state is 0 and the second state is 0.078 and the third one is 0.0079. The probability of transitioning into the second state is 1 and the probability of staying in the first and third are 0.

Step 4

At the end, IBM equity was selected with a number state of 2 due to its less difference. Continuous Markov process was used when the random variable of interest follows a continuous autoregressive process. The first-order autoregressive process (AR(1)) showed as below:

$$z_t = \rho z_{t-1} + \varepsilon_t$$

Where |p| < 1 and the error term is white noise with variance.

There are two methods which are Tauchen and Rouwenhorst methods. Based on the tauchen method, the Markov chain is separate into upper and lower bounds:

$$z_1 = -\lambda$$
$$z_N = \lambda$$

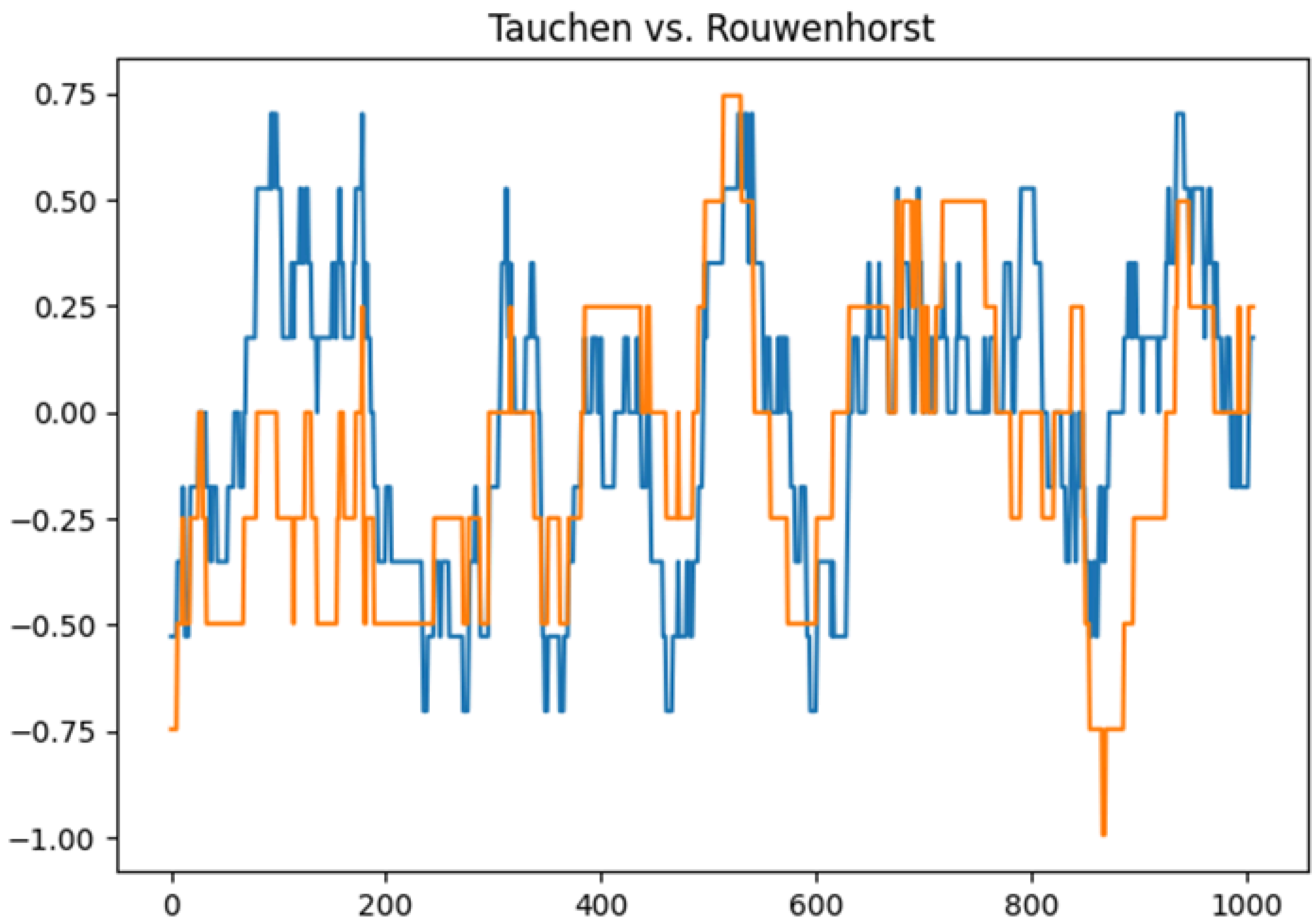
While the rouwenhorst method begins with equally spaced grid:

$$\mathcal{Z} = \{z_1, \ldots, z_N\} \text{ where } z_1 = -z_N.$$

First, it checks the parameters for the tauchen and rouwen method.

```
Checking the unconditional mean.... Tauchen Mean: [-7.70988212e-18] Rouwen Mean: [2.21659111e-18] Mean: 0
Checking the unconditional sd.... Tauchen sd: [0.35623664] Rouwen sd: [0.3537409] sd: 0.35102742508891926
```

The plot was generated as shown below where both movements are almost close to each other.




```
Real mean: 0 Tauchen mean: 0.0001932970402472001 Rouwen mean: -0.055766112360370955
Real sd: 0.35102742508891926 Tauchen sd: 0.33888113602688685 Rouwen_sd: 0.3461776763852721
Real rho: 0.975 Tauchen rho: 0.9630614657210393 Rouwen rho: 0.9747555843234116
```

Although there is a slight difference with the mean value from both methods, it still closes to zero. This scenario implied that the rouwenhorst method was slightly biased compared to the Tauchen method. Regarding the standard deviation, both methods have a close value with the real standard deviation.

The autocorrelation of the real value is 0.975, and the methods of Tauchen and Rouwenhorst are 0.96 and 0.97, respectively which are close enough to the real value. Thus, both methods can be applied to neutralize the biases and errors.

Reference

WorldQuant University, (2024). Reinforcement Learning: Intro To Markov Processes. WorldQuant University. Retrieved at: https://learn.wqu.edu/my-path/courses/stochastic-modeling/modules/m-4-reinforcement-learning-hidden-markov-models-for-detecting-regime-changes/tasks/lesson-1-reinforcement-learning-intro-to-markov-processes-lesson-notes