Capstone project

Quinn Le

A TIME SERIES APPROACH FOR OPTION PRICING

GOAL STATEMENT

The goal is to see if the time series approach yields a better result than Black-Scholes formula.

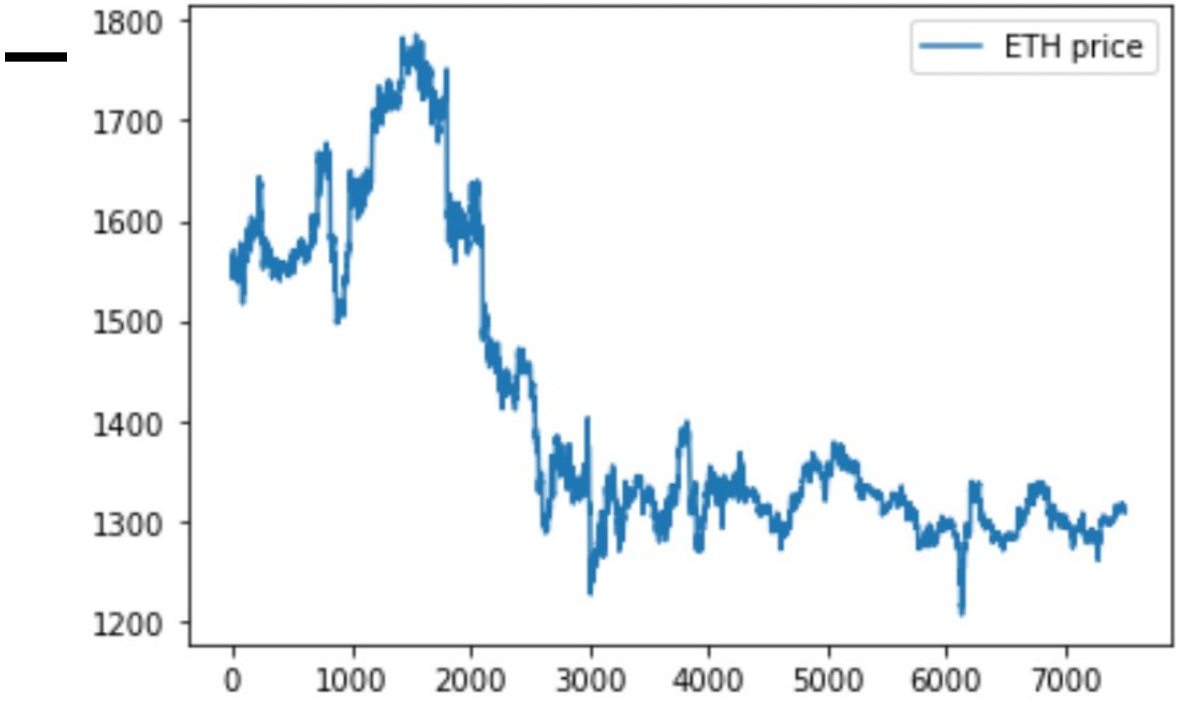
The option price is stochastic which is determined by supply and demand. Therefore, the price can be different from the price from Black-Scholes model (which is a well-known model in Finance which is used to price option). We would like to answer the question whether the time series models give a better prediction than the Black-Scholes model.

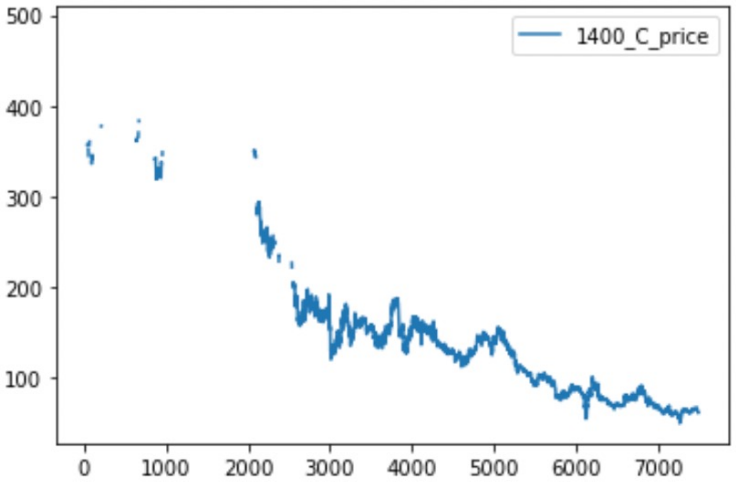
DATA OVERVIEW

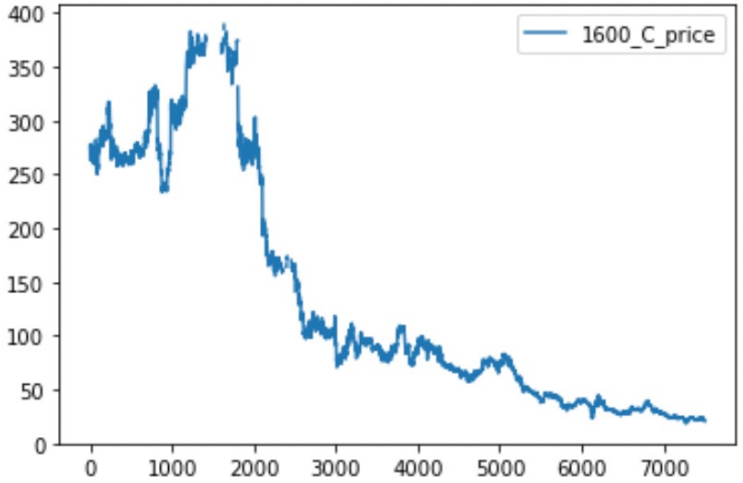
- Ethereum option data from Sep 2022 to Oct 2022
- The option type is European Call option

index	exchange	symbol	timestamp	local_timestamp	asks[0].price	asks[0].amount
0	deribit	ETH-3SEP22-1600-C	1662019222226000	1662019222230765	NaN	NaN
1	deribit	ETH-3SEP22-1600-C	1662019253509000	1662019253518236	NaN	NaN
2	deribit	ETH-3SEP22-1600-C	1662019266770000	1662019266780457	0.014	230.0
3	deribit	ETH-3SEP22-1600-C	1662019266777000	1662019266784004	0.014	230.0
4	deribit	ETH-3SEP22-1600-C	1662019273021000	1662019273026484	0.014	192.0

```
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   exchange         66607 non-null  object
 1   symbol           66607 non-null  object
 2   timestamp        66607 non-null  int64
 3   local_timestamp  66607 non-null  int64
 4   asks[0].price    66603 non-null  float64
 5   asks[0].amount   66603 non-null  float64
 6   bids[0].price    66607 non-null  float64
 7   bids[0].amount   66607 non-null  int64
 8   asks[1].price    66525 non-null  float64
 9   asks[1].amount   66525 non-null  float64
 10  bids[1].price    66606 non-null  float64
 11  bids[1].amount   66606 non-null  float64
 12  asks[2].price    65936 non-null  float64
 13  asks[2].amount   65936 non-null  float64
 14  bids[2].price    66597 non-null  float64
 15  bids[2].amount   66597 non-null  float64
 16  asks[3].price    63434 non-null  float64
 17  asks[3].amount   63434 non-null  float64
 18  bids[3].price    66475 non-null  float64
 19  bids[3].amount   66475 non-null  float64
 20  asks[4].price    52364 non-null  float64
 21  asks[4].amount   52364 non-null  float64
 22  bids[4].price    65986 non-null  float64
 23  bids[4].amount   65986 non-null  float64
dtypes: float64(19), int64(3), object(2)
```





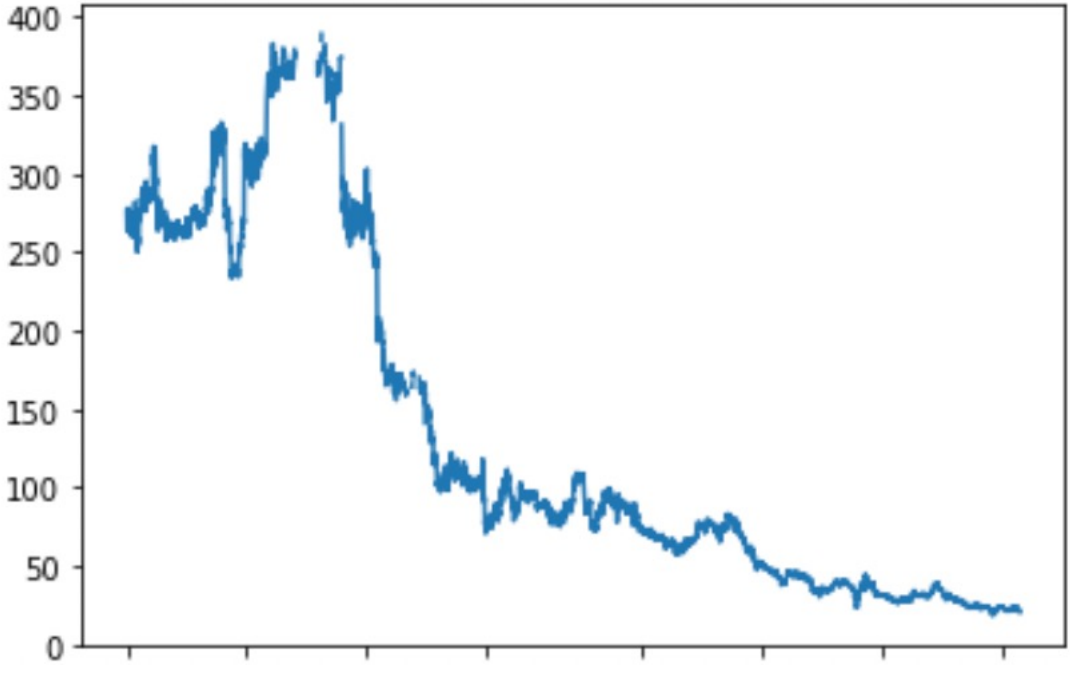


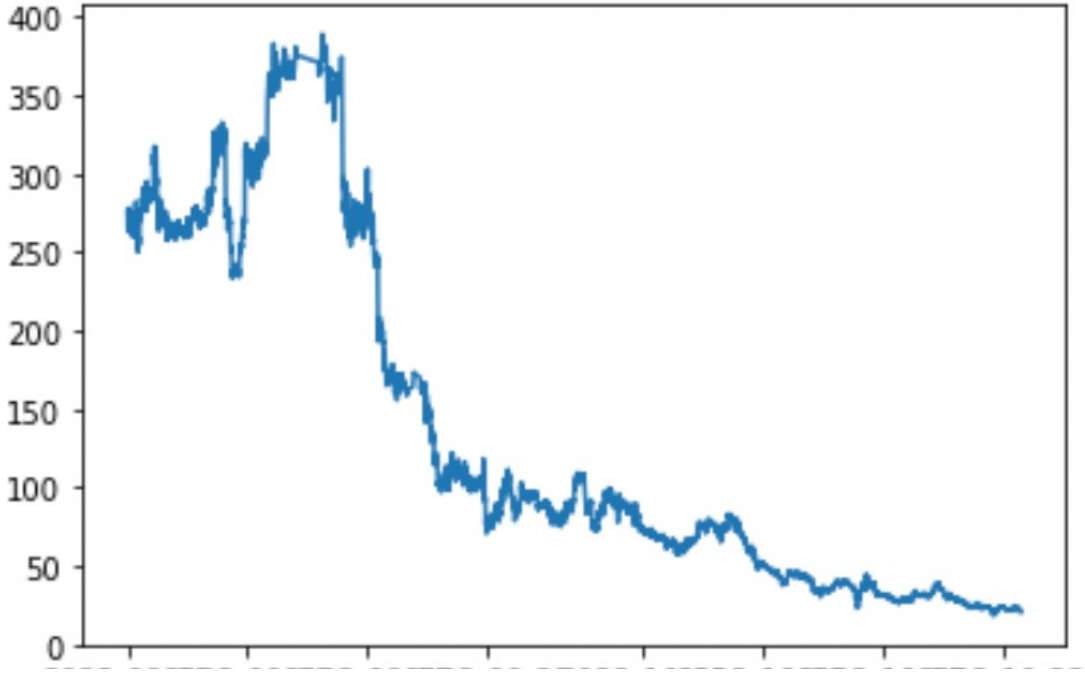
MISSING VALUE SUMMARY

```
index          0
timestamp      0
spot           3
1400        2848
1400_vol    2478
1500        1107
1500_vol     653
1600         445
1600_vol     225
1700         253
1700_vol      41
1800         305
1800_vol      39
1900         470
1900_vol      35
2000         669
2000_vol      58
dtype: int64
```

DATA WRANGLING

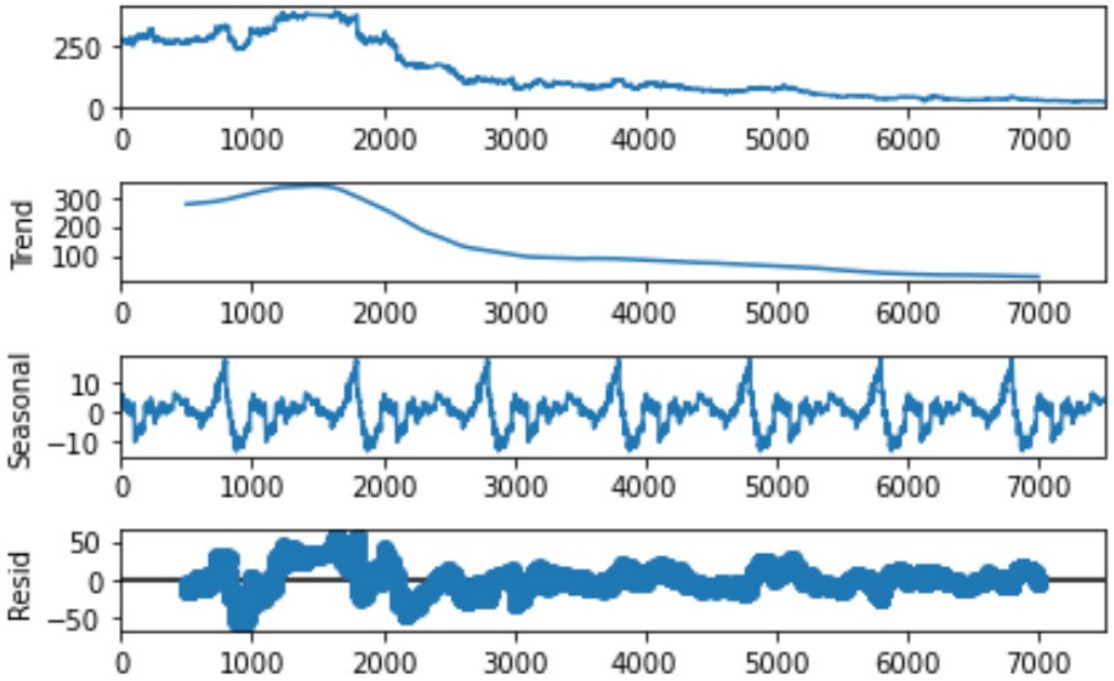
- Remove data points without bids[0].price or asks[0].price
- Remove data points with big bid-ask spread (> 0.005)
- Remove data points with low transaction volume (< 10 ETH)
- Fill NaN by linear interpolation



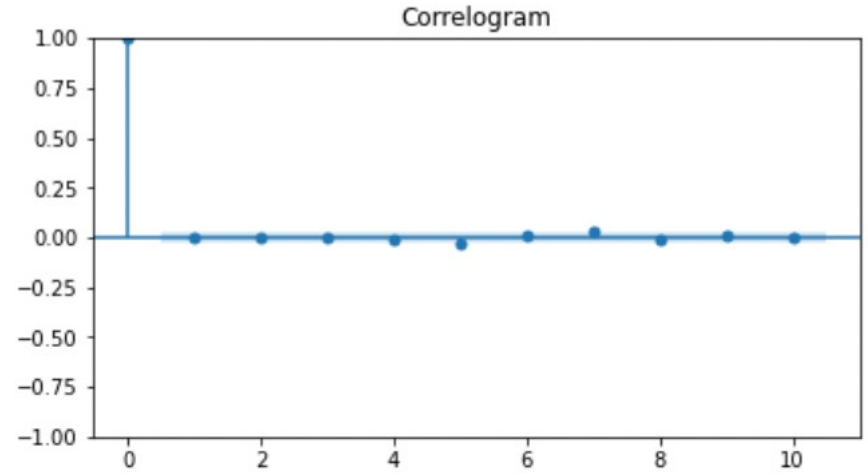


	timestamp	spot	1400	1400_vol	1500	1500_vol	1600	1600_vol	1700	1700_vol	1800	1800_vol
0	2022-09-01 00:00:00	1564.625			318.79234375000000	67534.0	276.1563125	56480.0	238.99646875	67628.0	206.53050000000000	98030.0
1	2022-09-01 00:10:00	1557.725			314.6604500000000	54743.0	272.60187500000000	36322.0	235.21647500000000	65656.0	203.2831125	63065.0
2	2022-09-01 00:20:00	1556.975			314.11970625	38877.0	272.08138125	47025.0	235.10322500000000	64104.0	202.79599375	59136.0
3	2022-09-01 00:30:00	1545.425			307.539575	47710.0	265.8131	68580.0	229.4956125	95785.0	197.8144	78468.0
4	2022-09-01 00:40:00	1541.975			304.92555625	100832.0	263.29223125	123147.0	227.44131250000000	157334.0	195.44533125000000	136925.0
5	2022-09-01 00:50:00	1550.125			309.24993750000000	44529.0	268.171625	79119.0	230.96862500000000	115110.0	199.57859375	109302.0
6	2022-09-01 01:00:00	1554.975			311.77248750000000	45685.0	270.17690625	62438.0	232.85750625000000	81340.0	200.98051875	98422.0
7	2022-09-01 01:10:00	1551.975			309.23101875	36268.0	268.10368125000000	70971.0	231.63226875000000	55266.0	199.81678125	65929.0
8	2022-09-01 01:20:00	1553.075			310.22673125	32326.0	268.68197500000000	45081.0	231.79644375	54625.0	199.95840625000000	72852.0
9	2022-09-01 01:30:00	1556.475			312.46235625	35020.0	270.82665000000000	61903.0	233.47125000000000	48688.0	201.17439375	55784.0
10	2022-09-01 01:40:00	1558.625			313.67328125	34718.0	271.59040625	101851.0	234.5730625	78432.0	201.8419375	93669.0
11	2022-09-01 01:50:00	1558.625			313.67328125	20142.0	271.20075	61881.0	234.5730625	55007.0	201.8419375	53036.0
12	2022-09-01 02:00:00	1564.475			317.58842500000000	27279.0	274.95648125000000	62842.0	237.8002	64181.0	204.946225	62524.0
13	2022-09-01 02:10:00	1568.325			319.9383	30168.0	276.8093625	64462.0	239.1695625	42078.0	206.62681875	50269.0
14	2022-09-01 02:20:00	1570.575			321.1825875	20709.0	277.991775	20733.0	240.297975	24382.0	207.31590000000000	21207.0
15	2022-09-01 02:30:00	1561.2350000000000			314.97916125000000	29378.0	272.82581625000000	19842.0	235.746485	22500.0	202.96055	39949.0
16	2022-09-01 02:40:00	1552.99			310.20975250000000	31781.0	268.2790225	26813.0	231.78375750000000	39413.0	199.17096750000000	42693.0
17	2022-09-01 02:50:00	1556.335			312.0451675	24758.0	270.0241225	17749.0	233.06116625000000	48520.0	200.76721500000000	32996.0
18	2022-09-01 03:00:00	1559.185			314.1757775	26722.0	272.07778250000000	34604.0	235.04713875	50382.0	202.30425375	42872.0
19	2022-09-01 03:10:00	1559.33			314.204995	20835.0	272.4929175	41884.0	234.67916500000000	51997.0	202.3230675	33122.0
20	2022-09-01 03:20:00	1554.075			311.20351875	28078.0	269.6320125	64424.0	232.3342125	63617.0	200.08715625000000	54173.0
21	2022-09-01 03:30:00	1553.375			310.28665625	21712.0	268.73387500000000	45753.0	232.22956250000000	55663.0	199.6086875	29426.0
22	2022-09-01 03:40:00	1546.025			306.11295	54549.0	264.75678125	133965.0	228.0386875	139286.0	196.73168125	110895.0
23	2022-09-01 03:50:00	1548.3600000000000			306.96237	40124.0	265.93083000000000	77682.0	229.54437000000000	96029.0	197.02881000000000	105490.0

ARIMA MODEL



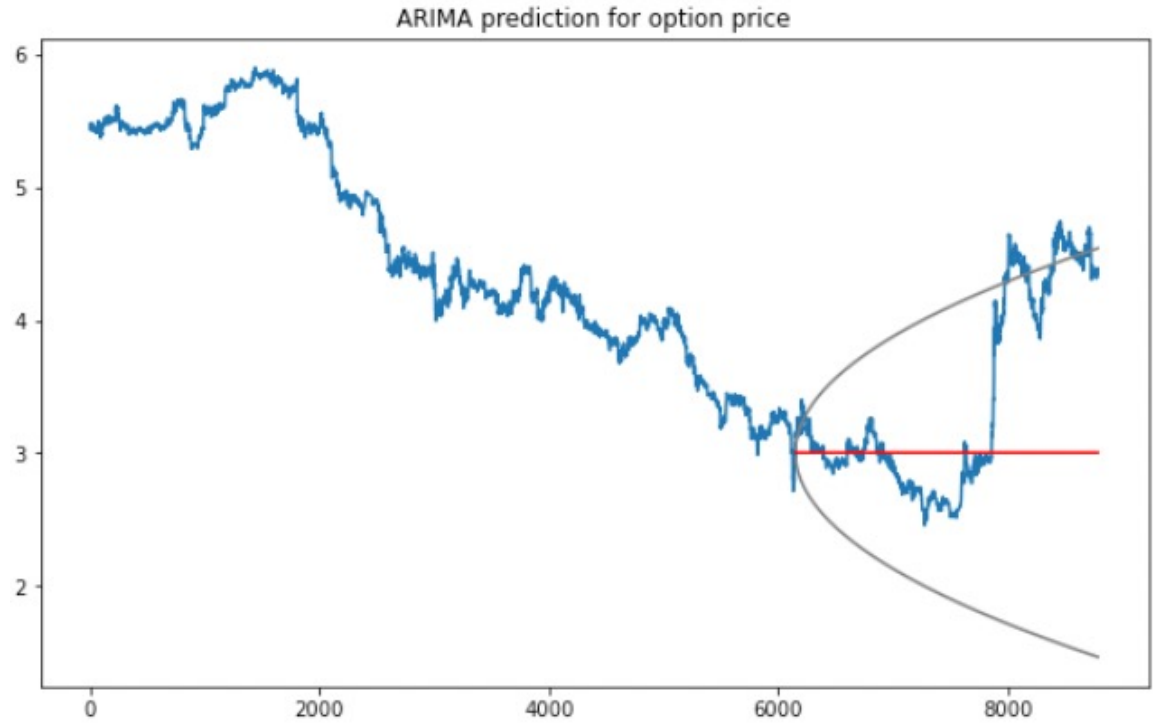
The best model is ARIMA(0,1,1)




```
 ARIMA(0,1,1)(0,0,0)[31]             : AIC=-35343.255, Time=0.24 sec

Best model:  ARIMA(0,1,1)(0,0,0)[31] intercept
Total fit time: 142.044 seconds
                                SARIMAX Results
==========================================================================
=====
Dep. Variable:                          y   No. Observations:
6372
Model:                SARIMAX(0, 1, 1)   Log Likelihood                1767
5.330
Date:                 Sun, 27 Nov 2022   AIC                          -3534
4.661
Time:                         16:47:24   BIC                          -3532
4.382
Sample:                              0   HQIC                         -3533
7.640
                                - 6372
Covariance Type:                  opg
```

ARIMA RESULT



- Mean squared error: 0.679
- Mean absolute error: 0.581

BLACK-SCHOLES FORMULA

$$C = N(d_1)S_t - N(d_2)Ke^{-rt}$$

$$\text{where } d_1 = \frac{\ln \frac{S_t}{K} + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}$$

$$\text{and } d_2 = d_1 - \sigma\sqrt{t}$$

C = call option price

N = CDF of the normal distribution

S_t = spot price of an asset

K = strike price

r = risk-free interest rate

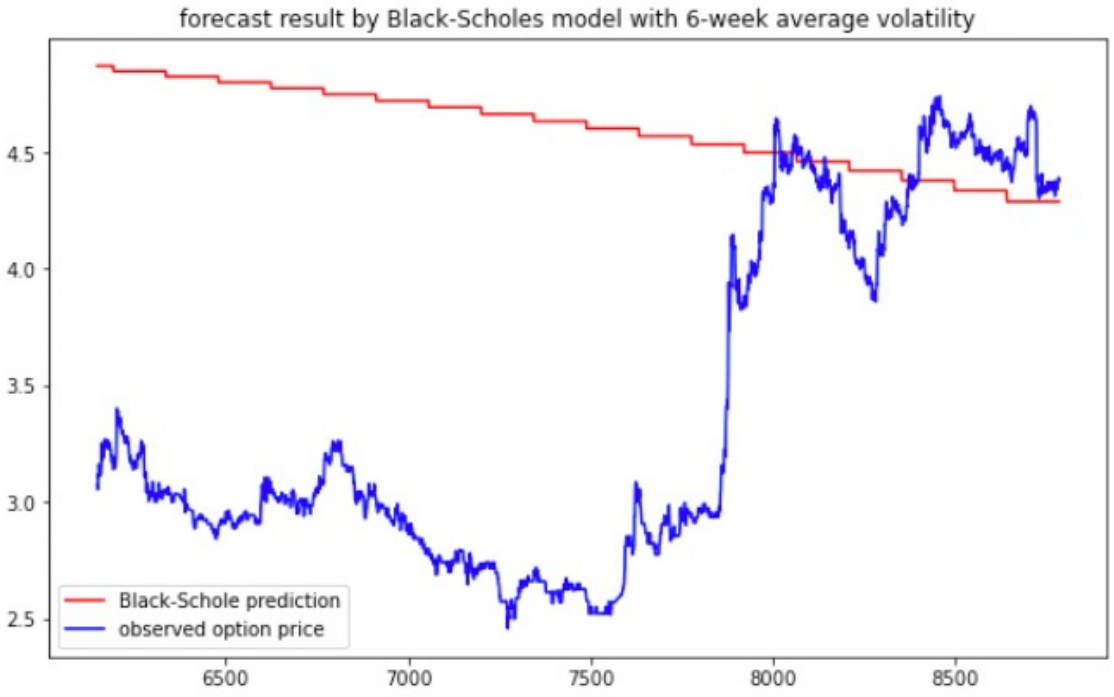
t = time to maturity

σ = volatility of the asset

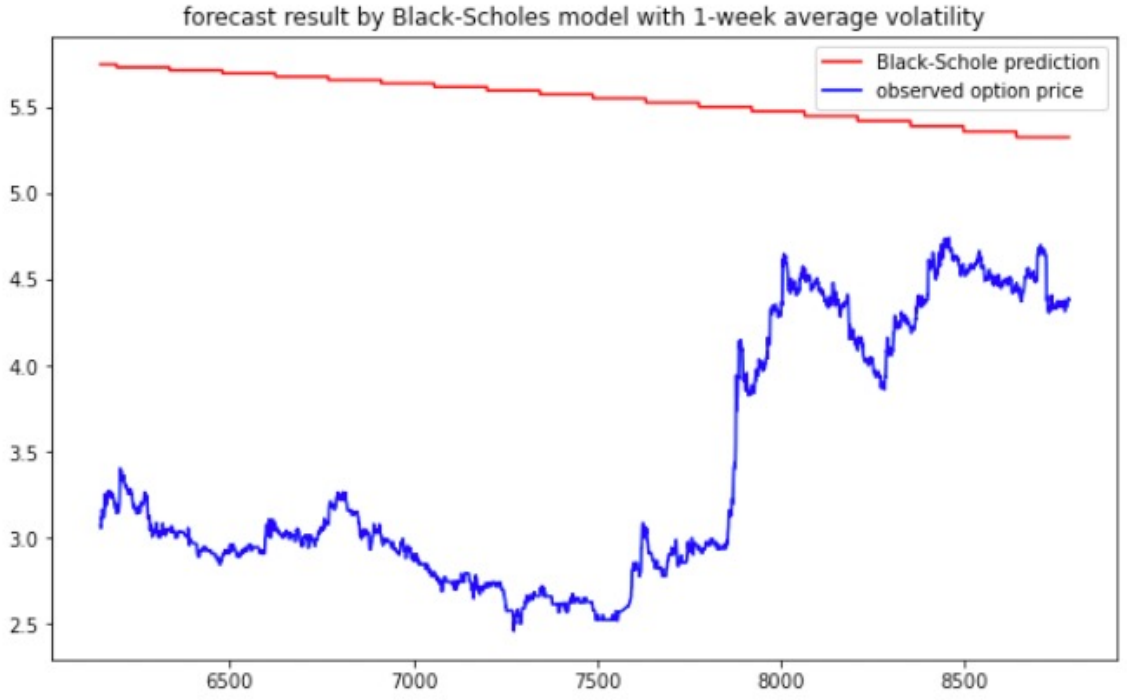
In our problem:

- Spot price = 1267.475
- Maturity time = Nov 25th 2022
- Strike price = 1700
- Volatility: 6-week average vol & 1-week average vol

BLACK-SCHOLES RESULTS



MSE: 2.214 MAE: 1.269



MSE: 5.343 MAE: 2.159

SUMMARY AND PROPOSED IMPROVEMENTS

- ARIMA: 0.679
- BS with 6-week volatility: 2.214
- BS with 1-week volatility: 5.343

Overall, ARIMA performs better in terms of MSE and MAE. Moreover, it gives a 95% confidence interval with covers most of the observed data.

The Black Scholes models with 6-week volatility performs better than the 1-week one in term of MSE. However, the price predicted by the later might be useful in some cases, for example, option seller.

PROPOSED IMPROVEMENTS

- For now, we simply use the average volatility for our model. Therefore, better volatility modeling will improve the accuracy for Black-Scholes model (GARCH for example).
- Neural network approach: Recurrent neural network (RNN) or Long short term memory LSTM etc.

THANK YOU FOR YOUR ATTENTION!