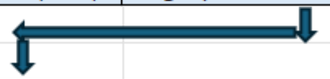
OPTIVER-TRADING AT THE CLOSE

DIVYAMAN SINGH RAWAT

Stock ID	Input sequence for target prediction at time ID K					Target prediction at time K
0	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 0 at K*
1	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 1 at K*
...	...	...	...	...	...	...
199	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 199 at K*

Stock ID	Input sequence for target prediction at time ID K					Target prediction at time K
0	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 0 at K*
1	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 1 at K*
...	...	...	...	...	...	...
199	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 199 at K*



Stock ID	Input sequence for target prediction at time ID (K+1)					Target prediction at time (K + 1)
0	*Target at (K - N + 1)*	*Target at (K - N + 2)*	...	*Target at (K - 1)*	***Target at K***	*Target prediction for stock ID 0 at (K+1)*
1	*Target at (K - N + 1)*	*Target at (K - N + 2)*	...	*Target at (K - 1)*	***Target at K***	*Target prediction for stock ID 1 at (K+1)*
...	...	...	...	...	...	...
199	*Target at (K - N + 1)*	*Target at (K - N + 2)*	...	*Target at (K - 1)*	***Target at K***	*Target prediction for stock ID 199 at (K+1)*

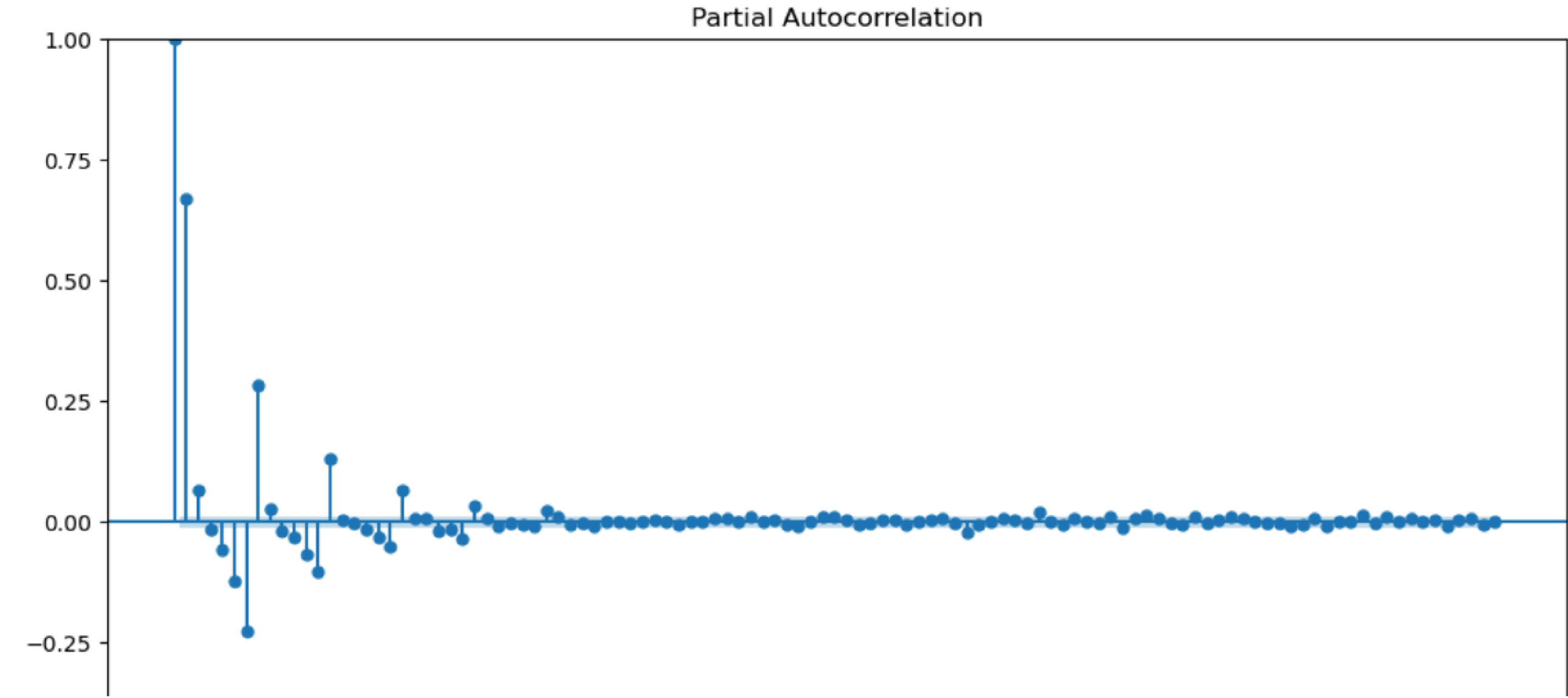
Stock ID	Input sequence for target prediction at time ID K					Target prediction at time K
0	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 0 at K*
1	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 1 at K*
...	...	...	...	...	...	...
199	*Target at (K - N)*	*Target at (K - N + 1)*	...	*Target at (K - 2)*	*Target at (K - 1)*	*Target prediction for stock ID 199 at K*

Stock ID	Input sequence for target prediction at time ID (K+1)					Target prediction at time (K + 1)
0	*Target at (K - N + 1)*	*Target at (K - N + 2)*	...	*Target at (K - 1)*	***Target at K***	*Target prediction for stock ID 0 at (K+1)*
1	*Target at (K - N + 1)*	*Target at (K - N + 2)*	...	*Target at (K - 1)*	***Target at K***	*Target prediction for stock ID 1 at (K+1)*
...	...	...	...	...	...	...
199	*Target at (K - N + 1)*	*Target at (K - N + 2)*	...	*Target at (K - 1)*	***Target at K***	*Target prediction for stock ID 199 at (K+1)*

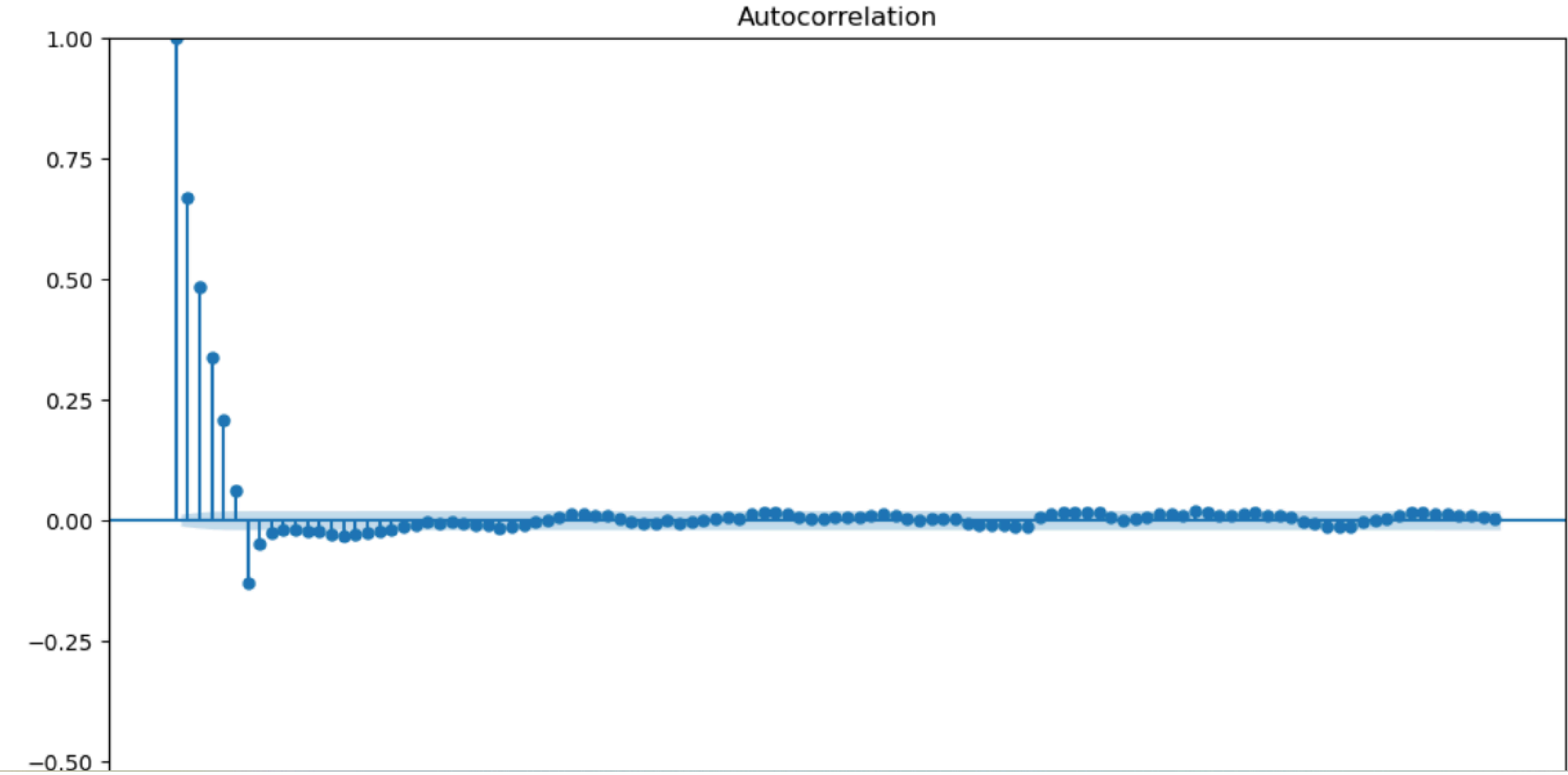
Stock ID	Input sequence for target prediction at time ID (K+2)					Target prediction at time (K + 2)
0	*Target at (K - N + 2)*	*Target at (K - N + 1)*	...	*Target at K*	***Target at (K+1)***	*Target prediction for stock ID 0 at (K+2)*
1	*Target at (K - N + 2)*	*Target at (K - N + 1)*	...	*Target at K*	***Target at (K+1)***	*Target prediction for stock ID 1 at (K+2)*
...	...	...	...	...	...	...
199	*Target at (K - N + 2)*	*Target at (K - N + 1)*	...	*Target at K*	***Target at (K+1)***	*Target prediction for stock ID 199 at (K+2)*

...

Visual inspection of PACF plot for stock ID 6: -



Visual inspection of ACF plot for stock ID 6: -



CHECK FOR STATIONARITY - 1

- **Condition 1**: -The mean must be constant and not vary with time.

CHECK FOR STATIONARITY - 2

- **Condition 1**: -The mean must be constant and not vary with time.
- **Condition 2**: - The variance must be constant and not vary with time.

CHECK FOR STATIONARITY - 3

- **Condition 1**: -The mean must be constant and not vary with time.
- **Condition 2**: - The variance must be constant and not vary with time.
- **Condition 3**: -There must be no periodicity in the data.

CHECK FOR STATIONARITY - 4

Check for conditions 1 and 2 (Constructing Samples)

stock_id	time_id	target
0	0	3.54
0	1	2.91
0	2	6.6
0	3	0.45
0	4	-3.4
0	5	-5.4
0	...	
0	...	
0	26452	-6.5
0	26453	2
0	26454	4.32

CHECK FOR STATIONARITY - 5

Check for conditions 1 and 2 (Constructing Samples)

stock_id	time_id	target
0	0	3.54
0	1	2.91
0	2	6.6
0	3	0.45
0	4	-3.4
0	5	-5.4
0	...	
0	...	
0	26452	-6.5
0	26453	2
0	26454	4.32

--- *Convert to samples* -->

Window size = 3

stock_id	time_id	target	
0	0	3.54	1st Sample
0	1	2.91	
0	2	6.6	
0	3	3.54	
0	4	2.91	2nd Sample
0	5	6.6	
0	...	...	
0	...	...	
0	26452	-6.5	8818th Sample
0	26453	2	
0	26454	4.32	

CHECK FOR STATIONARITY - 6

Check for conditions 1 and 2 (Constructing Samples)

stock_id	time_id	target
0	0	3.54
0	1	2.91
0	2	6.6
0	3	0.45
0	4	-3.4
0	5	-5.4
0	...	
0	...	
0	26452	-6.5
0	26453	2
0	26454	4.32

--- *Convert to samples* -->

Window size = 3

stock_id	time_id	target	
0	0	3.54	1st Sample
0	1	2.91	
0	2	6.6	
0	3	3.54	
0	4	2.91	2nd Sample
0	5	6.6	
0	...	...	
0	...	...	
0	26452	-6.5	8818th
0	26453	2	Sample
0	26454	4.32	

--- *Run ANOVA for the samples* --> *Run Levene test for the samples*

LSTM MODEL - 1

Constructing input sequence

Lagged Target Values with High DirectCorrelation: -			
- Lag 1			
- Lag 7			

LSTM MODEL - 2

Constructing input sequence

Lagged Target Values with High DirectCorrelation: -			
- Lag 1			
- Lag 7			

Lagged Target Values with High Indirect Correlation:			
- Lag 2			
- Lag 3			
- Lag 4			
- Lag 5			
- Lag 7			

LSTM MODEL - 3

Constructing input sequence

Lagged Target Values with High DirectCorrelation: -
- Lag 1
- Lag 7

Lagged Target Values with High Indirect Correlation:
- Lag 2
- Lag 3
- Lag 4
- Lag 5
- Lag 7

						Prediction for current Time ID	
						/\	
	Sequence					I	
----->	*Lag 7 Target*	*Lag 5 Target*	*Lag 4 Target*	*Lag 3 Target*	*Lag 2 Target*	*Lag 1 Target*	

LSTM MODEL - 4

Testing

Lagged Time IDs						time_id
26283	26285	26286	26287	26288	26289	26290

LSTM MODEL - 5

Testing

Lagged Time IDs						time_id
26283	26285	26286	26287	26288	26289	26290
26284	26286	26287	26288	26289	26290	26291
26285	26287	26288	26289	26290	26291	26292
26286	26288	26289	26290	26291	26292	26293
26287	26289	26290	26291	26292	26293	26294
26288	26290	26291	26292	26293	26294	26295
26289	26291	26292	26293	26294	26295	26296

LSTM MODEL - 6

Testing

Lagged Time IDs						time_id
26283	26285	26286	26287	26288	26289	26290
26284	26286	26287	26288	26289	26290	26291
26285	26287	26288	26289	26290	26291	26292
26286	26288	26289	26290	26291	26292	26293
26287	26289	26290	26291	26292	26293	26294
26288	26290	26291	26292	26293	26294	26295
26289	26291	26292	26293	26294	26295	26296
26290	26292	26293	26294	26295	26296	26297
26291	26293	26294	26295	26296	26297	26298
...	...	...	...	...	...	

OBSERVATIONS & CONCLUSIONS - 1

- The MAD achieved on the test data is an improvement over the MAD achieved by the winning team on the Kaggle competition.
- Further experimentation needed:-
  - Alternative models
  - Additional features

OBSERVATIONS & CONCLUSIONS - 2

- The MAD achieved on the test data is an improvement over the MAD achieved by the winning team on the Kaggle competition.
- Further experimentation needed:-
  - Alternative models
  - Additional features

Version 2 of the project planned soon.

CONTACT ME

E-mail divyamans@gmail.com

Linked-In https://www.linkedin.com/in/divyaman-rawat/

Project Github: - https://github.com/DivNewBeg/Optiver-trading-at-close

THANK YOU!