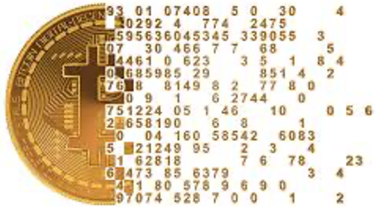


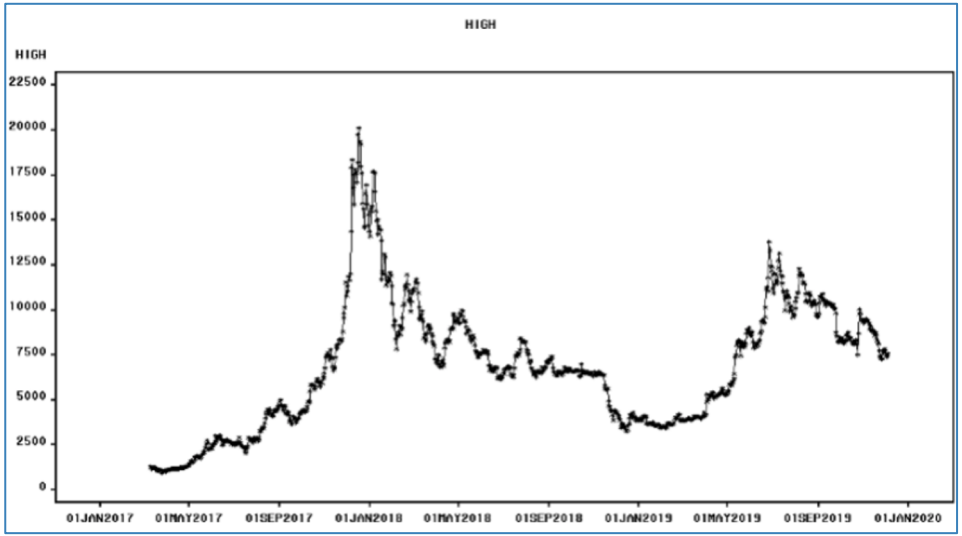
Modeling Bitcoin Price

Yichen Li
Junyi Qian
Junzhe Yin
Meizi Yu
Shengqi Zhou

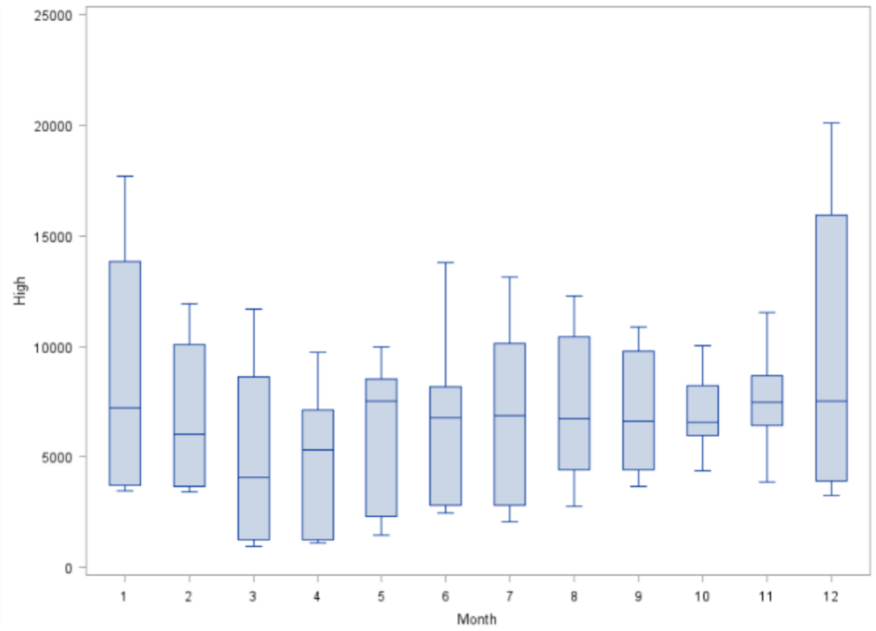
1 Introduction: Bitcoin

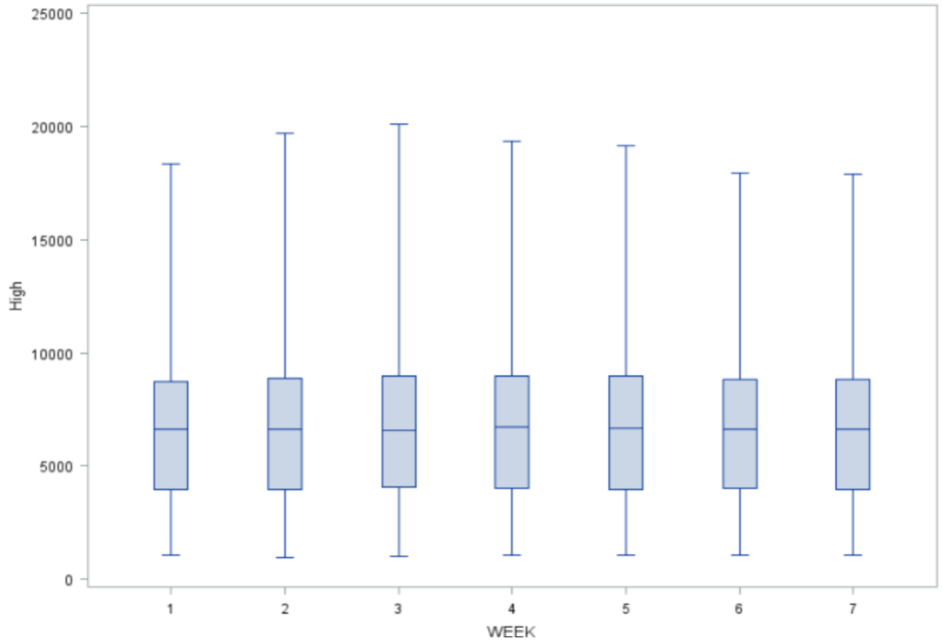
1000 observations, 30 or 40 hold out samples

- Date : from Mar 10, 2017 to Dec 04, 2019
- High : highest recorded trading price of the day
- Volume : the monetary value of the currency traded in a 24 hour period, denoted in USD



Box Plot


25000
20000
15000
10000
5000
0
High
1
2
3
4
5
6
7
8
9
10
11
12
Month


25000
20000
15000
10000
5000
0
High
1
2
3
4
5
6
7
WEEK

Periodogram table of *High*

- Top 11 P-value：lag 1, 2, 3, 4, 5, 6, 7, 9, 10, 13, 16
- Linear trend+sine+cosine
- MAPE(40)<MAPE(30)
- 30 hold-out sample is better

The SAS System

Obs	FREQ	PERIOD	P_01
1	0.00000	.	0.00
2	0.00628	1000.00	2349.67
3	0.01257	500.00	4641.64
4	0.01885	333.33	1385.61
5	0.02513	250.00	608.94
6	0.03142	200.00	315.21
7	0.03770	166.67	229.10
8	0.04398	142.86	330.45
9	0.05027	125.00	61.56
10	0.05655	111.11	197.91
11	0.06283	100.00	112.34
12	0.06912	90.91	13.14
13	0.07540	83.33	23.87
14	0.08168	76.92	119.66
15	0.08796	71.43	56.57
16	0.09425	66.67	63.85
17	0.10053	62.50	124.54

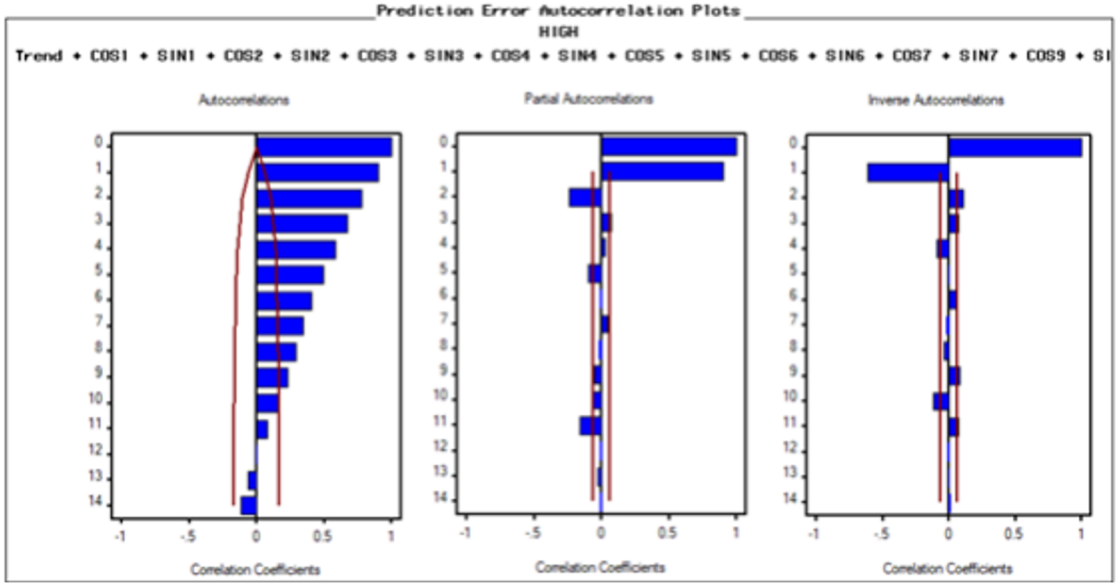
Forecast Model	Model Title	Mean Absolute Percent Error
☑	Linear Trend + COS1 + SIN1 + COS2 + SIN2 + COS3 + SIN3 + COS4 + SIN4 + COS5 + SI	2.10944
☐	Linear Trend + COS1 + SIN1 + COS2 + SIN2 + COS3 + SIN3 + COS4 + SIN4 + COS5 + SI	11.79660
☐	Linear Trend + COS1 + SIN1 + COS2 + SIN2 + COS3 + SIN3 + COS4 + SIN4 + COS5 + SI	16.80105

30 hold-out sample+AR(2)
30 hold-out sample
40 hold-out sample

2 30 hold-out Sample Cyclical Model

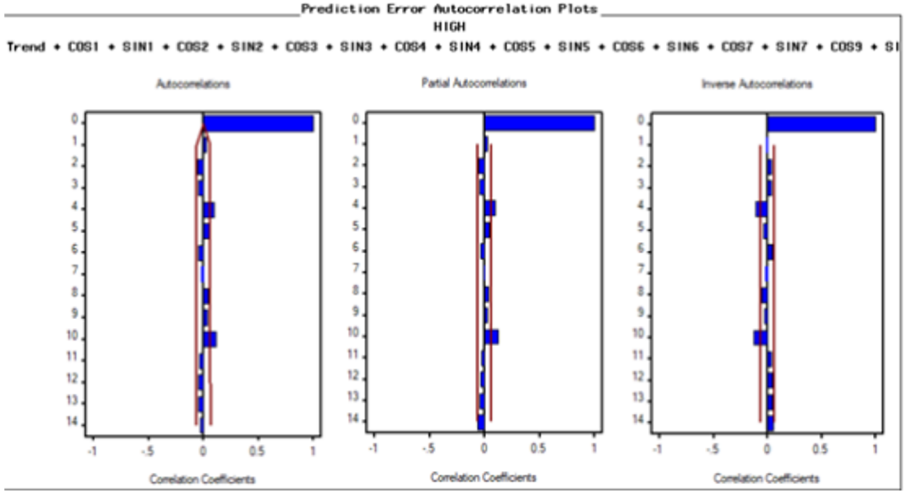
30 hold-out sample:

- The ACF decaying slowing
- non-stationary
- PACF Chopped off at lag 2
- AR(2) model
- MAPE of AR(2) is Lowest
- Cyclical model + AR(2) best



Cyclic model:
Linear tread+cos1+sin1+cos2+sin2+...+cos7+sin7+cos9+sin9+cos10+sin10+cos13+sin13+cos16+sin16

Cyclical model + AR(2)



Parameter estimated &Function

Parameter Estimates

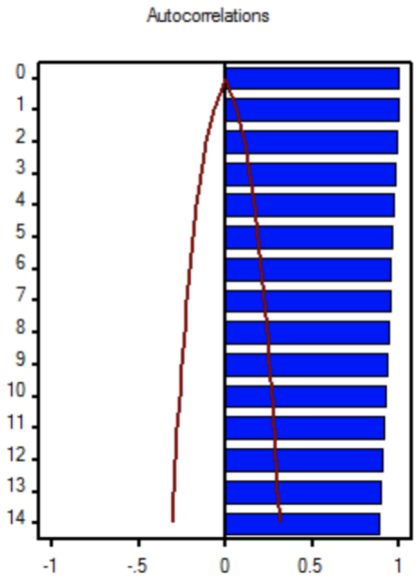
HIGH

+ SIN2 + COS3 + SIN3 + COS4 + SIN4 + COS5 + SIN5 + COS6 + SIN6 + COS7 + SIN7 + COS9 + SIN9 + COS10 + SIN10 + COS1

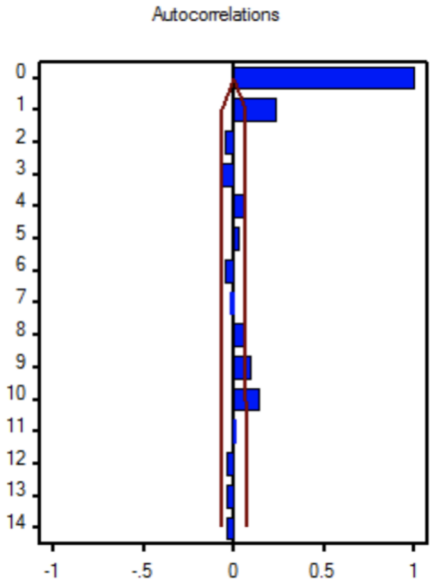
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Autoregressive, Lag 1	1.13109	0.0315	35.9208	<.0001
Autoregressive, Lag 2	-0.24948	0.0315	-7.9174	0.0014
Linear Trend	4.47061	1.9600	2.2809	0.0847
COS1	-908.74958	140.1897	-6.4823	0.0029
SIN1	2001	632.4732	3.1631	0.0341
COS2	-636.23996	139.6652	-4.5555	0.0104
SIN2	-2955	328.8228	-8.9855	0.0008
COS3	109.17799	138.8051	0.7866	0.4755
SIN3	-1659	232.6539	-7.1288	0.0020
COS4	-271.18061	137.6297	-1.9704	0.1201
SIN4	1057	187.8604	5.6280	0.0049
COS5	-742.58008	136.1663	-5.4535	0.0055
SIN5	179.43327	163.1929	1.0995	0.3333
COS6	19.10612	134.4485	0.1421	0.8939
SIN6	-693.00440	148.2505	-4.6745	0.0095
COS7	835.01409	132.5145	6.3013	0.0032
SIN7	-43.51802	138.6015	-0.3140	0.7692
COS9	-422.55900	128.1653	-3.2970	0.0300
SIN9	-479.29221	127.4029	-3.7620	0.0197
COS10	328.87902	125.8363	2.6135	0.0592
SIN10	-381.55353	123.9437	-3.0784	0.0370
COS13	-68.00717	118.7087	-0.5729	0.5974
SIN13	-518.09199	117.0237	-4.4272	0.0114
COS16	-287.57873	111.9815	-2.5681	0.0621
SIN16	-449.16938	111.7401	-4.0198	0.0159
Model Variance (sigma squared)	108365	.	.	.

3 ARIMA Model

- Non Stationary
- Difference Stationary TS
- ARIMA(p,d,q)
  - d : After dth differenced series is stationary
  - p : PACF chopped off after Lag p
  - q : ACF chopped off after Lag q

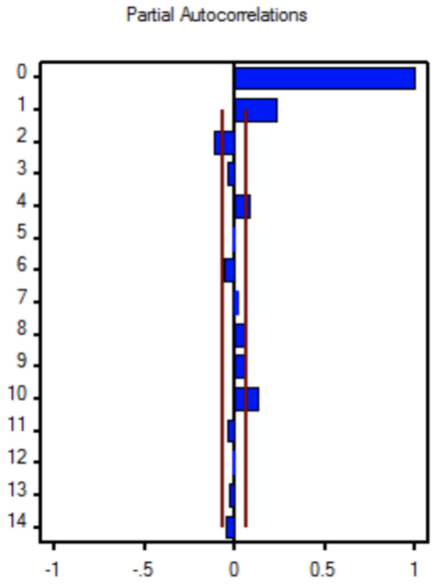


ACF of HIGH



ACF after first differencing

q=1



PACF after first differencing

p= 2 or 1

ARIMA（1，1，0）

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Autoregressive, Lag 1	0.25333	0.0315	8.0470	<.0001
Autoregressive, Lag 2	-0.10946	0.0315	-3.4771	0.0005
Model Variance (sigma squared)	113498	.	.	.

Check Intercept

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	8.50761	14.2525	0.5969	0.5554
Autoregressive, Lag 1	0.22738	0.0313	7.2613	<.0001
Model Variance (sigma squared)	117577	.	.	.

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Autoregressive, Lag 1	0.22783	0.0313	7.2801	<.0001
Model Variance (sigma squared)	117499	.	.	.

Check Statistic of Fit

Statistic of Fit	Value
Mean Square Error	26530.9
Root Mean Square Error	162.88315
Mean Absolute Percent Error	1.49682
Mean Absolute Error	123.32323
R-Square	0.948

$$\hat{P}_t = P_{t-1} + 0.228\ (P_{t-1} - P_{t-2})$$

ARIMA（0，1，1）

Check Intercept

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	8.50702	13.7348	0.6194	0.5407
Moving Average, Lag 1	-0.25190	0.0311	-8.0929	<.0001
Model Variance (sigma squared)	116694	.	.	.

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	-0.25218	0.0311	-8.1068	<.0001
Model Variance (sigma squared)	116619	.	.	.

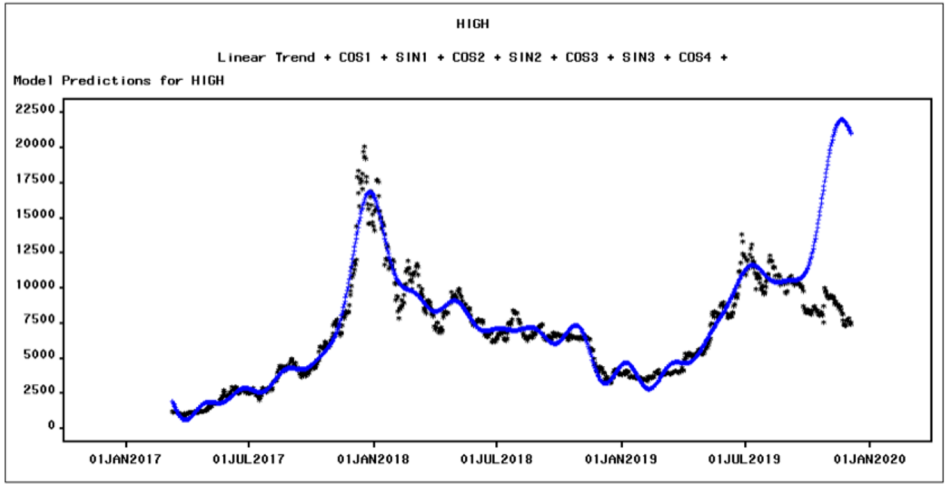
Check Statistic of Fit

Statistic of Fit	Value
Mean Square Error	26475.4
Root Mean Square Error	162.71273
Mean Absolute Percent Error	1.50846
Mean Absolute Error	124.35233
R-Square	0.949

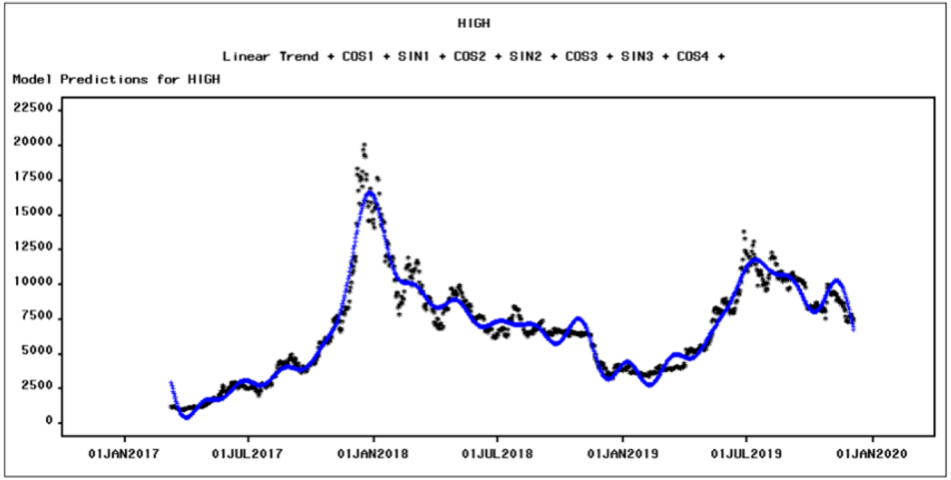
$$\hat{P}_t = P_{t-1} + 0.252\varepsilon_{t-1}$$

Where $\varepsilon_{t-1} = P_{t-1} - \hat{P}_{t-2}$

4 Decision of hold out sample



85 hold out sample



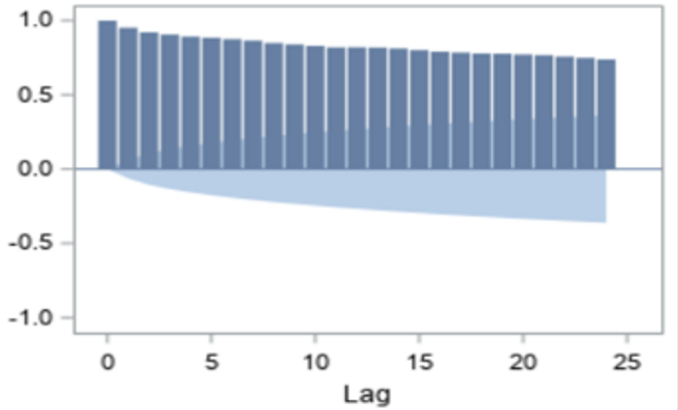
30 hold out sample

Models Comparison

Hold Out	Models	MAPE	MAE	RMSE	R-Square
30	Cyclical+AR(2)	2.10944	174.94778	201.75939	0.921
30	IAR(1,1)	1.49682	123.32323	162.88315	0.948
	IMA(1,1)	1.50846	124.35233	162.71273	0.949
15	IAR(1,1)	1.54067	117.84266	161.91157	0.620
	IMA(1,1)	1.55198	118.71474	162.39881	0.618

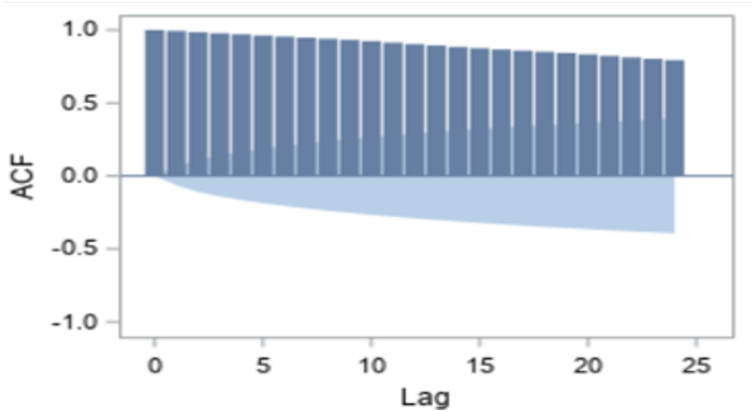
5 Transfer Function Model

- Relationship between High and Volume
- Based on the CCF find the parameters of TF
- Check residuals and make TF noise model



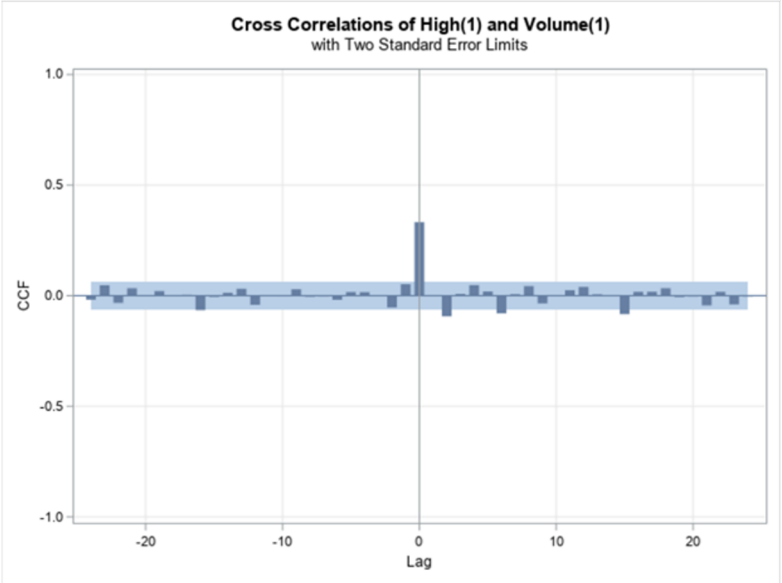
ACF of Volume

Volume as X



ACF of High

High as Y

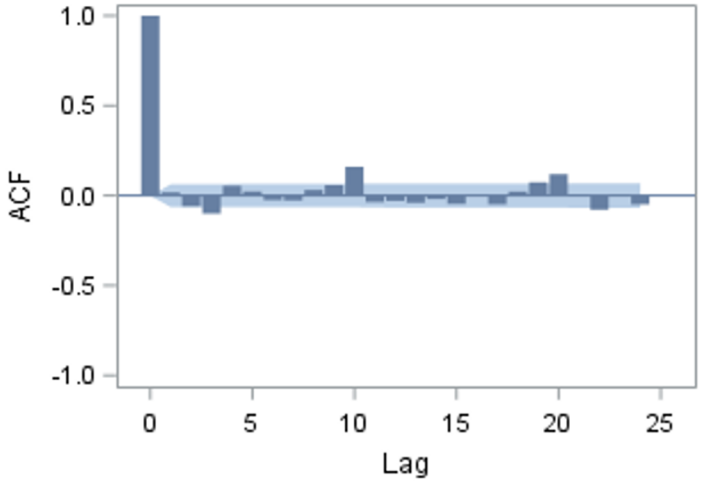


CCF of High(1) and Volume(1)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	62.51	6	<.0001	0.229	-0.018	-0.068	0.043	0.032	-0.044
12	98.15	12	<.0001	-0.024	0.056	0.092	0.147	0.013	-0.036
18	110.69	18	<.0001	-0.051	-0.039	-0.071	-0.028	-0.033	0.037
24	141.78	24	<.0001	0.091	0.121	0.006	-0.070	-0.024	-0.047
30	150.64	30	<.0001	-0.004	-0.031	-0.010	-0.020	0.014	0.083
36	159.70	36	<.0001	-0.033	-0.060	-0.032	-0.015	-0.002	-0.053
42	168.31	42	<.0001	-0.021	0.035	0.061	0.018	-0.028	-0.043
48	197.75	48	<.0001	0.083	0.123	0.046	-0.019	-0.060	-0.004

Crosscorrelation Check of Residuals with Input Volume									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	8.67	6	0.1930	-0.000	-0.008	0.032	0.016	-0.085	0.008
11	13.65	12	0.3237	0.045	-0.030	-0.004	0.019	0.041	-0.001
17	21.33	18	0.2630	0.000	-0.077	0.024	0.016	0.030	-0.006
23	24.77	24	0.4185	-0.007	-0.043	0.018	-0.034	-0.005	-0.009
29	32.47	30	0.3461	0.020	0.008	-0.069	0.004	0.050	-0.006
35	37.36	36	0.4064	-0.039	-0.043	0.027	0.008	-0.027	0.010
41	49.40	42	0.2015	-0.001	0.052	0.000	0.077	-0.058	0.009
47	53.53	48	0.2706	0.045	0.020	-0.006	0.034	-0.001	-0.023

Check of Residuals



ACF of Residuals
after MA(1) Noise model

VOLUME + IMA(1,1) NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	-0.22735	0.0315	-7.2245	<.0001
VOLUME	5.73711E-8	4.6923E-9	12.2267	<.0001
Model Variance (sigma squared)	101049	.	.	.

Parameters estimation of TF model

Thanks!

Any questions ?