Forecast Pressure Time Series in Tiantan, Beijing

Puvit Pracharktam 6031830321

1 Select attributes

Used Columns :

‘Year’

‘Month’

‘Date’

‘Time’

‘PRES’

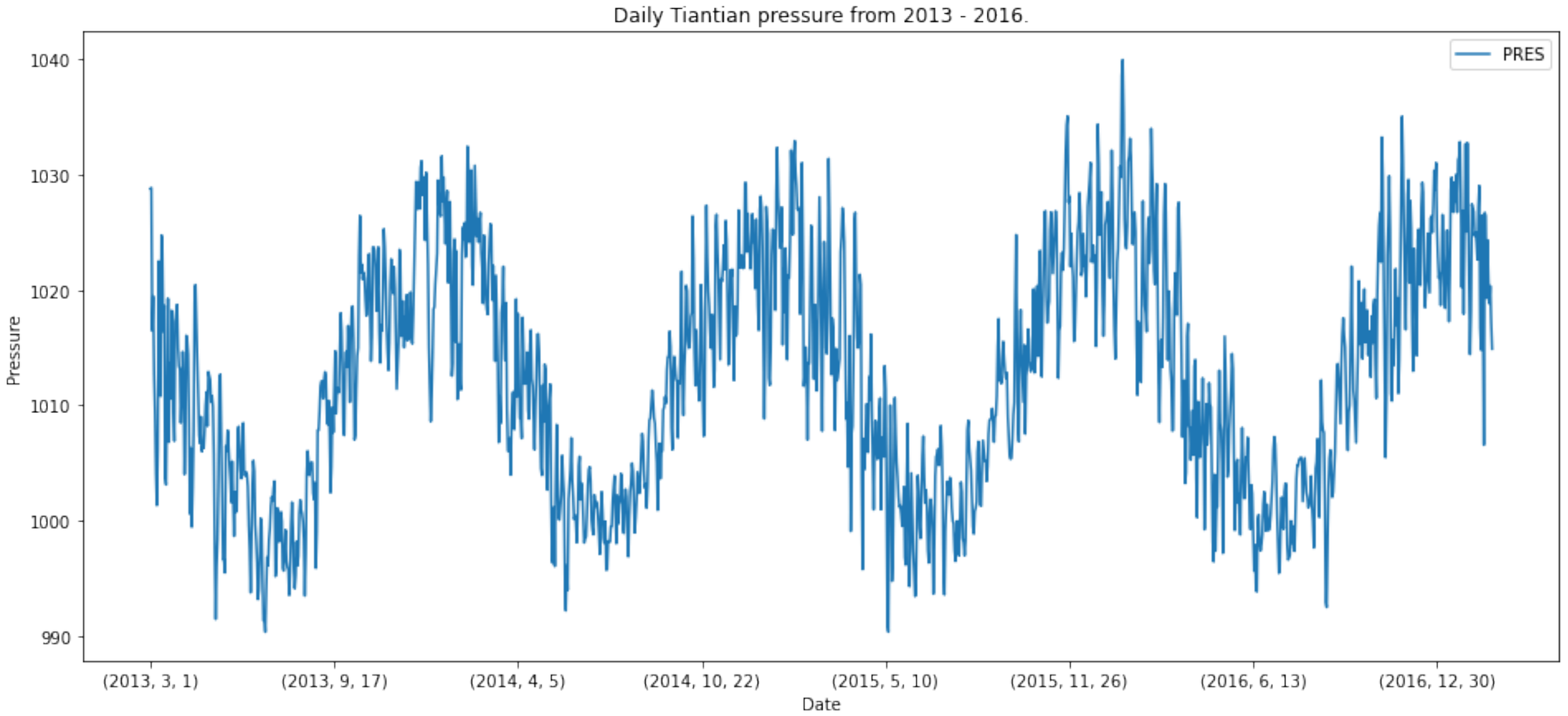
Group by

	No	year	month	day	hour	PM2.5	PM10	SO2	NO2	CO	O3	TEMP	PRES	DEWP	RAIN	wd	WSPM	station
0	1	2013	3	1	0	6.0	6.0	4.0	8.0	300.0	81.0	-0.5	1024.5	-21.4	0.0	NNW	5.7	Tiantan
1	2	2013	3	1	1	6.0	29.0	5.0	9.0	300.0	80.0	-0.7	1025.1	-22.1	0.0	NW	3.9	Tiantan
2	3	2013	3	1	2	6.0	6.0	4.0	12.0	300.0	75.0	-1.2	1025.3	-24.6	0.0	NNW	5.3	Tiantan
3	4	2013	3	1	3	6.0	6.0	4.0	12.0	300.0	74.0	-1.4	1026.2	-25.5	0.0	N	4.9	Tiantan
4	5	2013	3	1	4	5.0	5.0	7.0	15.0	400.0	70.0	-1.9	1027.1	-24.5	0.0	NNW	3.2	Tiantan

			PRES
year	**month**	**day**	
2013	**3**	**1**	1028.783333
		2	1028.850000
		3	1016.458333
		4	1019.454167
		5	1012.704167

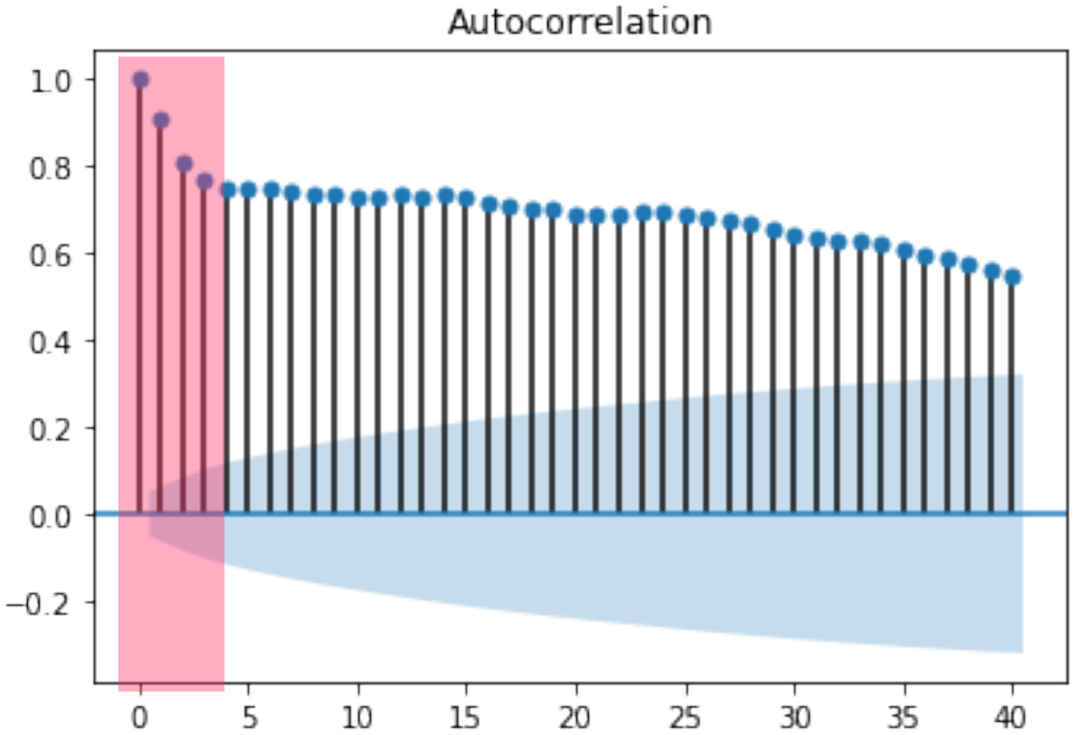
2 Aggregate pressure in each date by mean

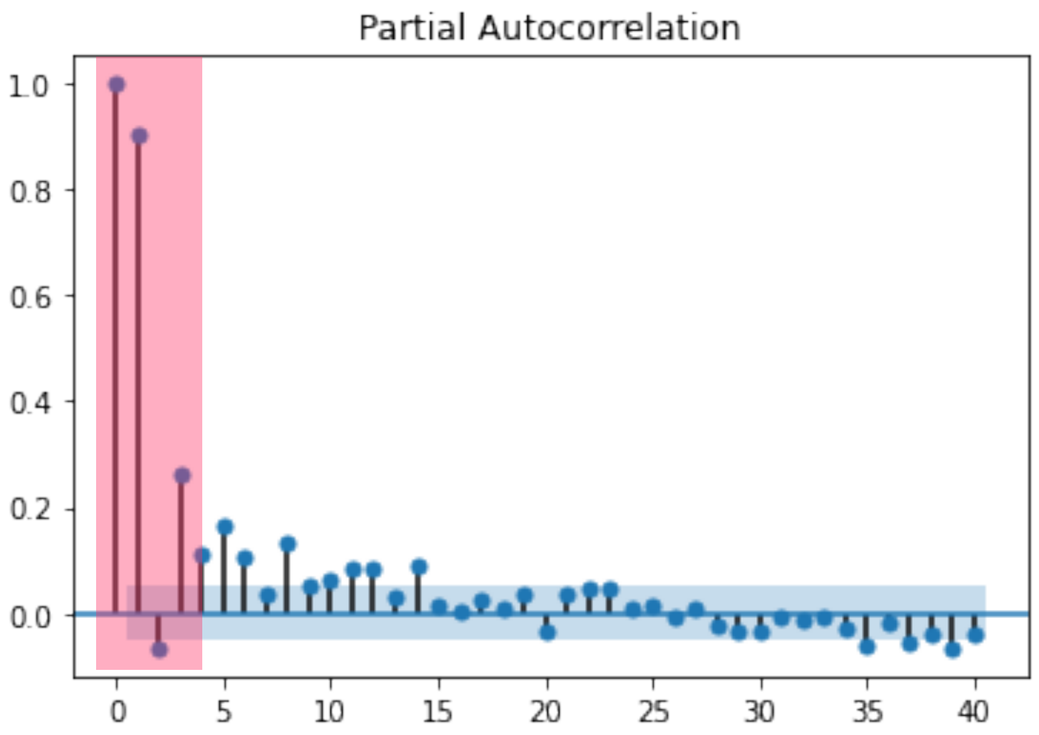
3 Visualize data (Trend, Seasonal)



There is same seasonal in every year

4 Calculate ACF and PACF





This graph can be interpreted that there is 4 lags of both ACF and PACF before stable

So, we must try ARIMA(4,1,4) first

5A Apply ARIMA model without deseasonal

ARIMA (4,1,1)

ARIMA Model Results

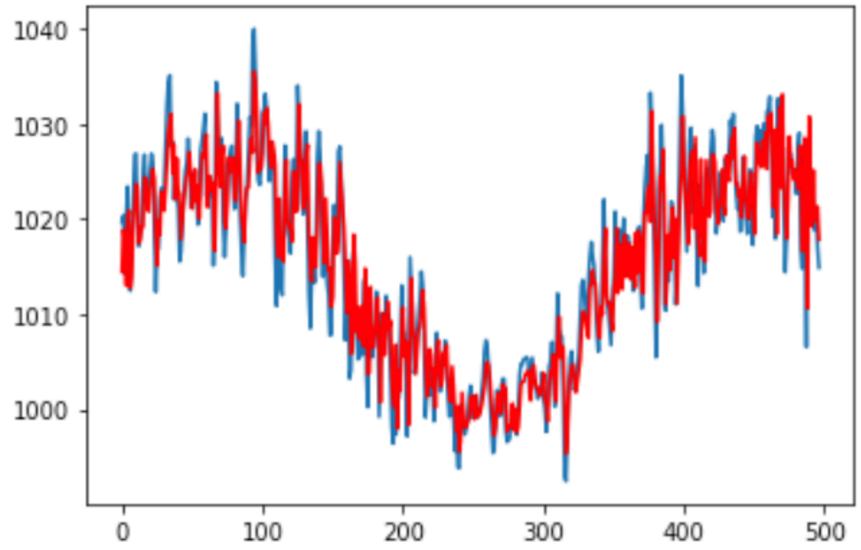
Dep. Variable:	D.PRES	No. Observations:	1460
Model:	ARIMA(4, 1, 1)	Log Likelihood	-4074.744
Method:	css-mle	S.D. of innovations	3.942
Date:	Mon, 09 Mar 2020	AIC	8163.489
Time:	23:35:09	BIC	8200.492
Sample:	1	HQIC	8177.293

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0018	0.025	0.071	0.944	-0.047	0.050
ar.L1.D.PRES	0.7538	0.032	23.321	0.000	0.690	0.817
ar.L2.D.PRES	-0.3228	0.033	-9.825	0.000	-0.387	-0.258
ar.L3.D.PRES	0.0890	0.033	2.681	0.007	0.024	0.154
ar.L4.D.PRES	-0.0506	0.028	-1.782	0.075	-0.106	0.005
ma.L1.D.PRES	-0.8731	0.020	-44.662	0.000	-0.911	-0.835

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.4251	-1.0261j	1.7561	-0.0993
AR.2	1.4251	+1.0261j	1.7561	0.0993
AR.3	-0.5459	-2.4714j	2.5310	-0.2846
AR.4	-0.5459	+2.4714j	2.5310	0.2846
MA.1	1.1454	+0.0000j	1.1454	0.0000

Test MSE: 16.675



5A Apply ARIMA model without deseasonal

ARIMA (4,1,0)

ARIMA Model Results

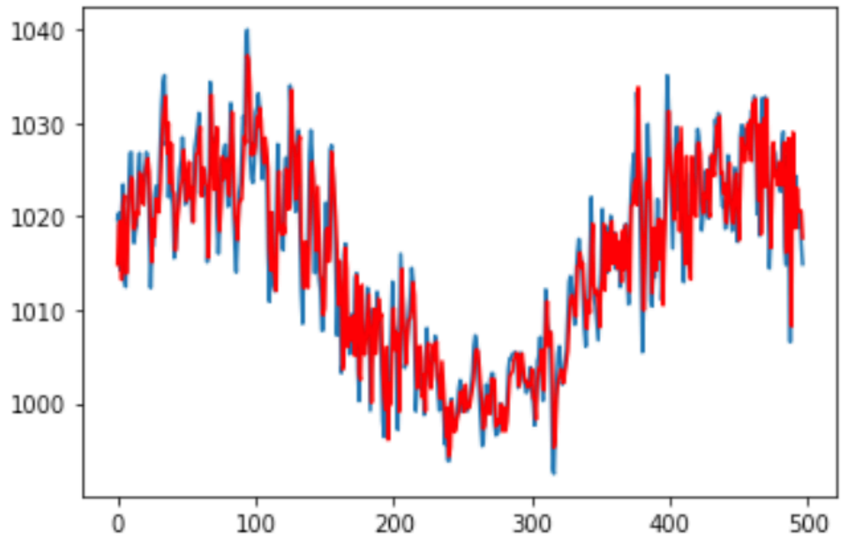
Dep. Variable:	D.PRES	No. Observations:	1460
Model:	ARIMA(4, 1, 0)	Log Likelihood	-4116.393
Method:	css-mle	S.D. of innovations	4.057
Date:	Tue, 10 Mar 2020	AIC	8244.787
Time:	12:46:30	BIC	8276.504
Sample:	1	HQIC	8256.619

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0047	0.062	-0.076	0.940	-0.126	0.117
ar.L1.D.PRES	-0.0467	0.026	-1.813	0.070	-0.097	0.004
ar.L2.D.PRES	-0.3439	0.026	-13.429	0.000	-0.394	-0.294
ar.L3.D.PRES	-0.1440	0.026	-5.632	0.000	-0.194	-0.094
ar.L4.D.PRES	-0.1785	0.026	-6.904	0.000	-0.229	-0.128

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	0.6740	-1.2163j	1.3906	-0.1695
AR.2	0.6740	+1.2163j	1.3906	0.1695
AR.3	-1.0772	-1.3176j	1.7019	-0.3591
AR.4	-1.0772	+1.3176j	1.7019	0.3591

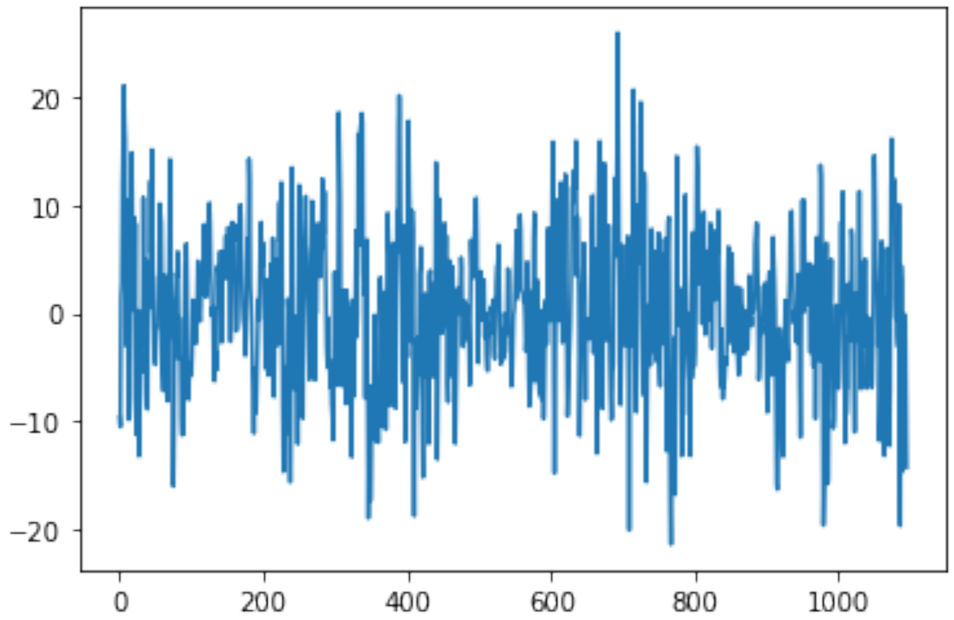
Test MSE: 17.858



5A Apply ARIMA model with deseasonal

Diff = Pres $_{\text{day I}}$ - Pres $_{\text{day i+365}}$

			diff
year	month	day	
2014	3	1	-9.612500
		2	-10.550000
		3	1.425000
		4	2.779167
		5	11.612500
...	...	...	...
2017	2	24	-7.308333
		25	-2.900000
		26	-0.158333
		27	-8.875000
		28	-14.308333



5B Apply ARIMA model with deseasonal

ARIMA (4,1,1)

ARIMA Model Results

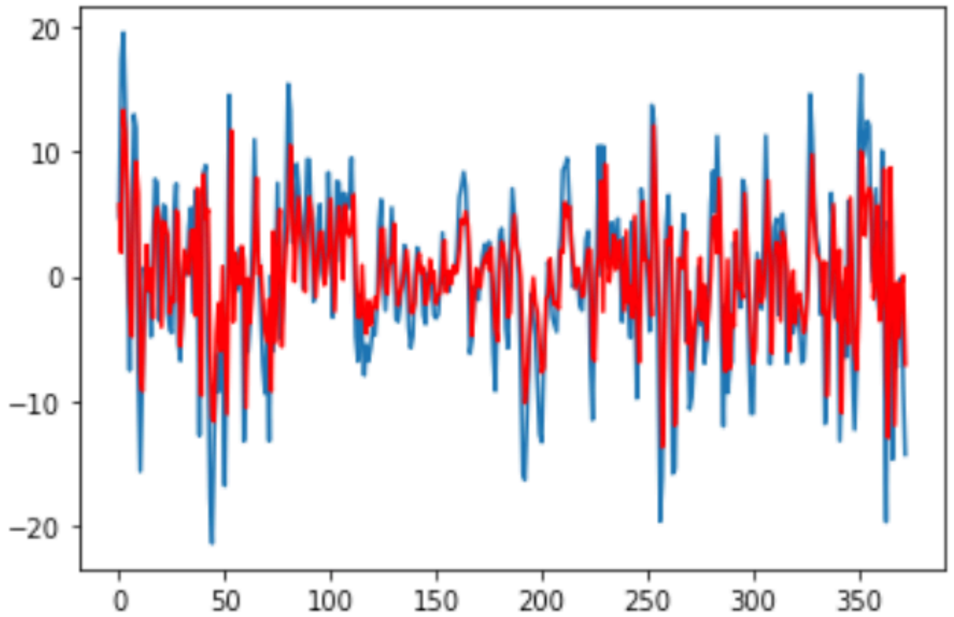
Dep. Variable:	D.diff	No. Observations:	1095
Model:	ARIMA(4, 1, 1)	Log Likelihood	-3392.629
Method:	css-mle	S.D. of innovations	5.347
Date:	Tue, 10 Mar 2020	AIC	6799.259
Time:	12:56:45	BIC	6834.248
Sample:	1	HQIC	6812.498

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0011	0.001	-1.015	0.310	-0.003	0.001
ar.L1.D.diff	0.7690	0.030	25.433	0.000	0.710	0.828
ar.L2.D.diff	-0.3011	0.038	-7.917	0.000	-0.376	-0.227
ar.L3.D.diff	0.0946	0.038	2.485	0.013	0.020	0.169
ar.L4.D.diff	-0.0256	0.030	-0.846	0.398	-0.085	0.034
ma.L1.D.diff	-0.9999	0.003	-350.585	0.000	-1.006	-0.994

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.9830	-1.0009j	2.2213	-0.0744
AR.2	1.9830	+1.0009j	2.2213	0.0744
AR.3	-0.1377	-2.8085j	2.8119	-0.2578
AR.4	-0.1377	+2.8085j	2.8119	0.2578
MA.1	1.0001	+0.0000j	1.0001	0.0000

Test MSE: 29.747



5B Apply ARIMA model with deseasonal

ARIMA (4,1,0)

ARIMA Model Results

Dep. Variable:	D.diff	No. Observations:	1095
Model:	ARIMA(4, 1, 0)	Log Likelihood	-3458.151
Method:	css-mle	S.D. of innovations	5.692
Date:	Tue, 10 Mar 2020	AIC	6928.302
Time:	12:57:46	BIC	6958.293
Sample:	1	HQIC	6939.650

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0050	0.095	-0.052	0.958	-0.191	0.182
ar.L1.D.diff	-0.0850	0.030	-2.870	0.004	-0.143	-0.027
ar.L2.D.diff	-0.3608	0.029	-12.273	0.000	-0.418	-0.303
ar.L3.D.diff	-0.1623	0.029	-5.530	0.000	-0.220	-0.105
ar.L4.D.diff	-0.2013	0.030	-6.779	0.000	-0.260	-0.143

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	0.6504	-1.1963j	1.3617	-0.1707
AR.2	0.6504	+1.1963j	1.3617	0.1707
AR.3	-1.0536	-1.2527j	1.6368	-0.3613
AR.4	-1.0536	+1.2527j	1.6368	0.3613

Test MSE: 30.060