Time Series and Business Forecasting

To –
Prof. Yamini Ma’am

By –
Nitish Sou

About Dataset

- Dongsi is a residential district in Dongcheng, Beijing
- Population :8,22,000 (as of 2018)
- Our dataset (source: UCI) contains the data relating to air quality from March 1st, 2013 to February 28th, 2017
- Total No of observations: 35064

Description of Dataset

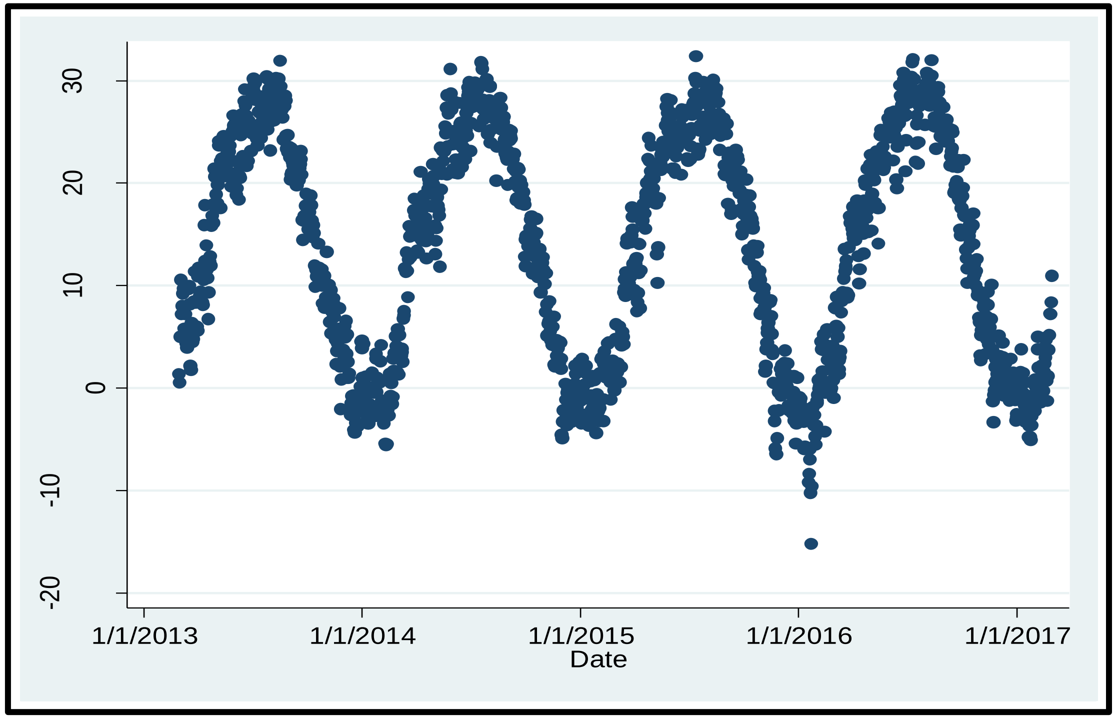
No: row number
year: year of data in this row
month: month of data in this row
day: day of data in this row
hour: hour of data in this row
PM2.5: PM2.5 concentration (ug/m^3) (**Independent**)
PM10: PM10 concentration (ug/m^3) (**Independent**)
SO2: SO2 concentration (ug/m^3) (**Independent**)
NO2: NO2 concentration (ug/m^3) (**Independent**)
CO: CO concentration (ug/m^3) (**Independent**)
O3: O3 concentration (ug/m^3) (**Independent**)
TEMP: temperature (degree Celsius) (**Dependent**)

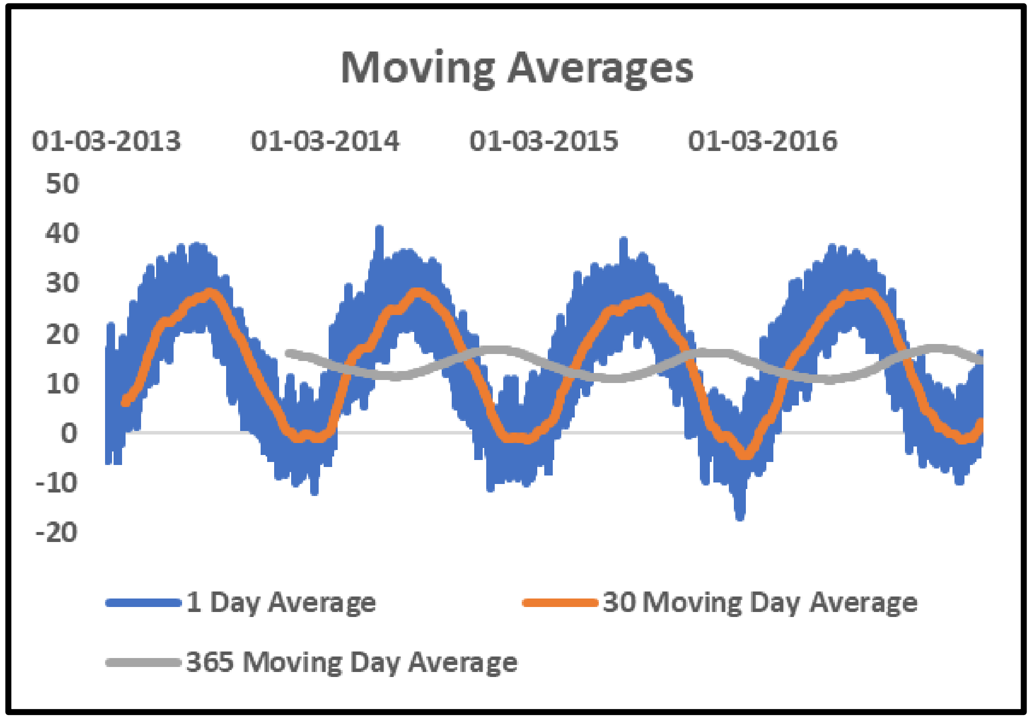
```
  obs:        35,064
 vars:            29                          16 Aug 2021 18:18
 size:     4,803,768
```

variable name	storage type	display format	value label	variable label
no	long	%8.0g		No
year	int	%8.0g		
month	byte	%8.0g		
day	byte	%8.0g		
hour	byte	%8.0g		
pm25	str5	%9s		PM2.5
pm10	str5	%9s		PM10
so2	str8	%9s		SO2
no2	str8	%9s		NO2
co	str5	%9s		CO
o3	str8	%9s		O3
temp	str12	%12s		TEMP
pres	str11	%11s		PRES
dewp	str5	%9s		DEWP
rain	str4	%9s		RAIN
wd	str3	%9s		
wspm	str4	%9s		WSPM
station	str6	%9s		
pm25_n	float	%9.0g		
pm10_n	float	%9.0g		
so2_n	float	%9.0g		
no2_n	float	%9.0g		
co_n	float	%9.0g		
o3_n	float	%9.0g		
temp_n	float	%9.0g		
pres_n	float	%9.0g		
dewp_n	float	%9.0g		
rain_n	float	%9.0g		
wspm_n	float	%9.0g		

Temp

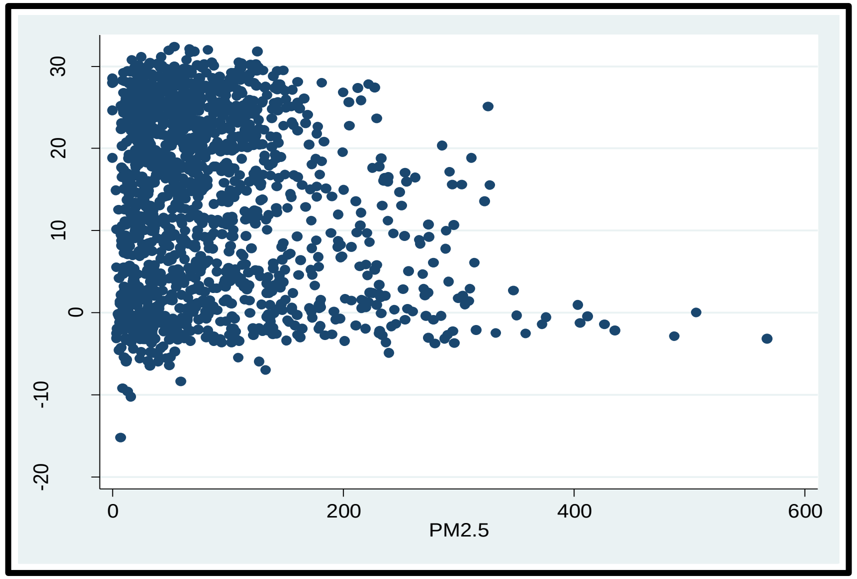
Observation: Temperature is showing seasonal variation with time.

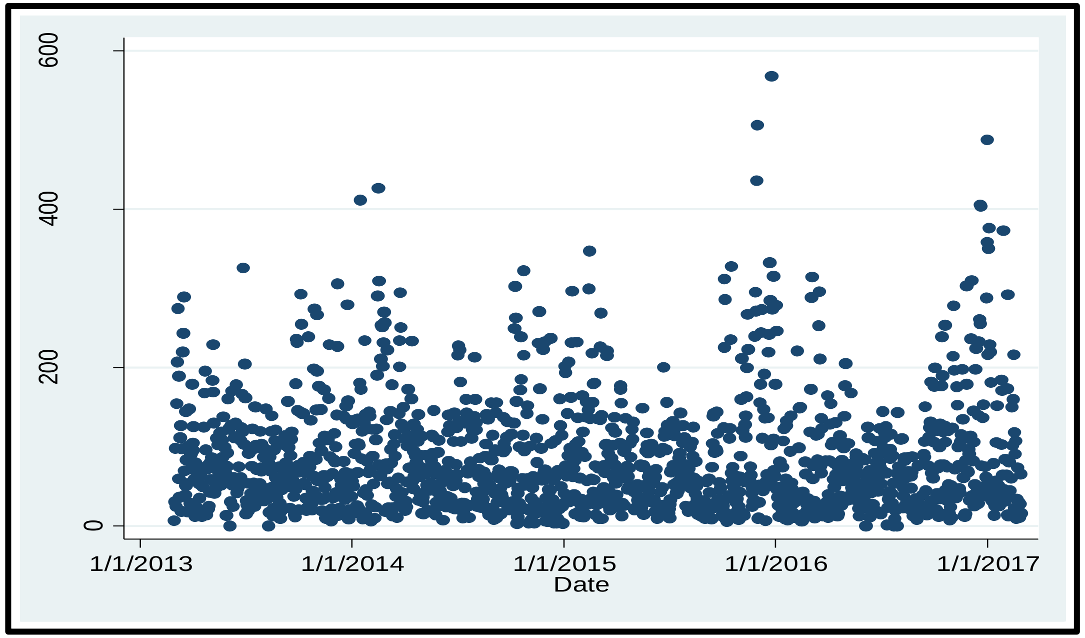


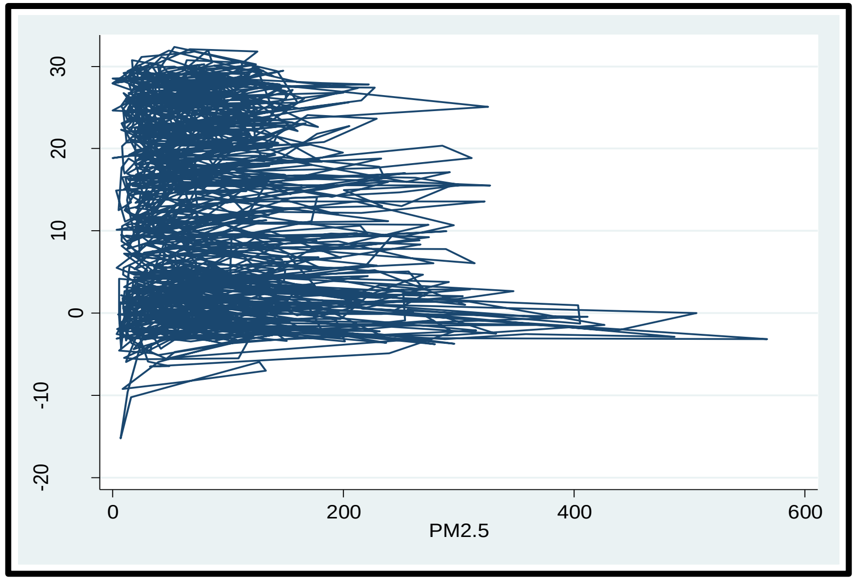


Temp vs PM2.5

Observation: As PM2.5 rises, temperature falls. This shows that the relationship between PM2.5 and Temperature is negatively correlated.

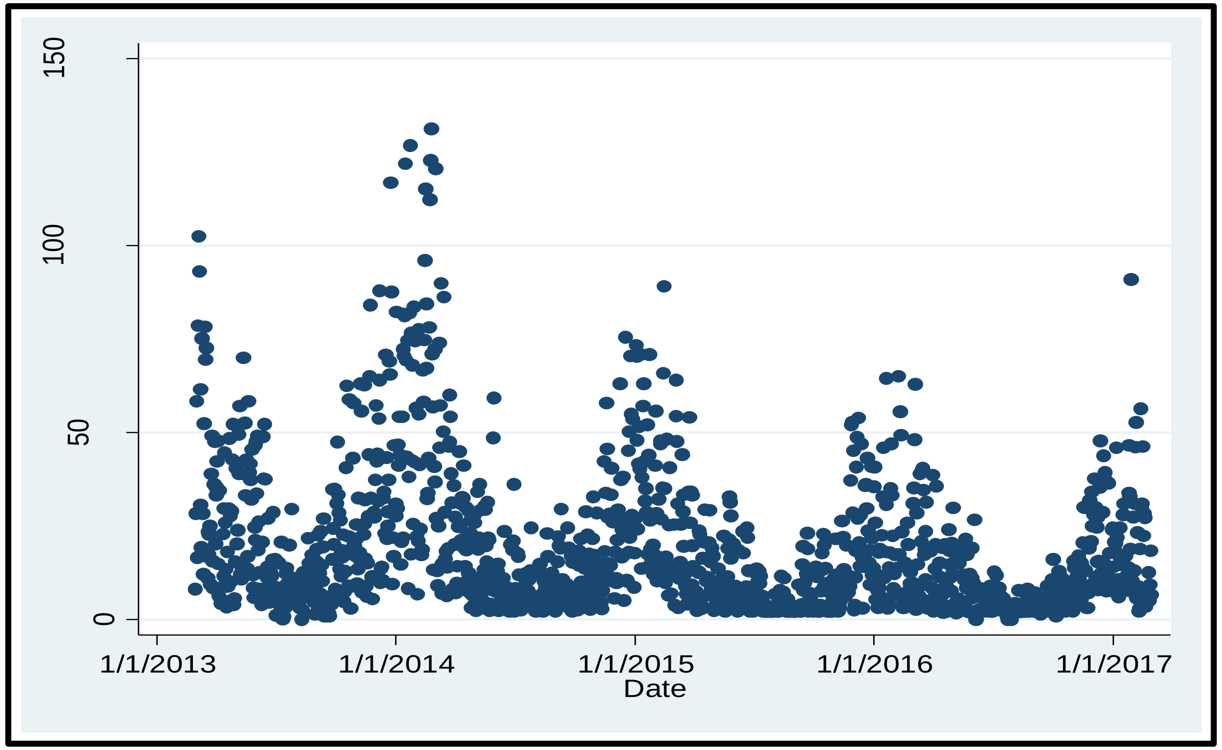


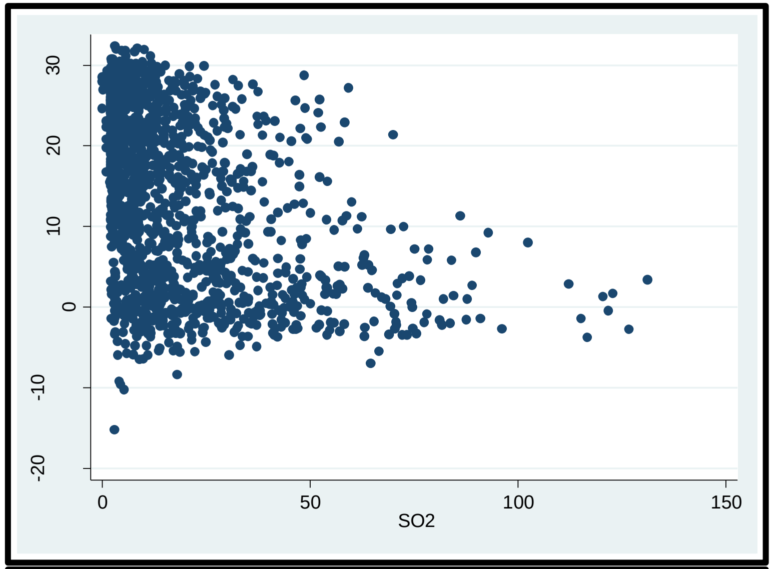


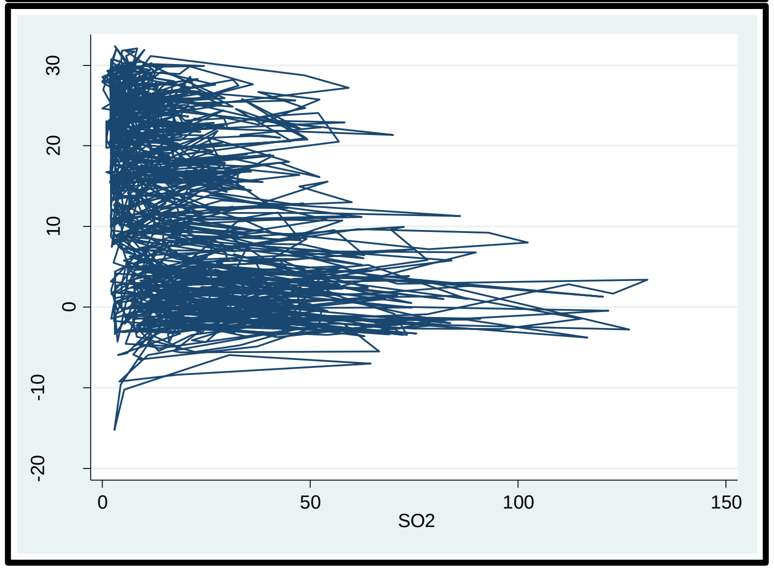


Temp vs SO2

Observation: As SO2 rises, temperature falls. This shows that the relationship between SO2 and Temperature is negatively correlated.

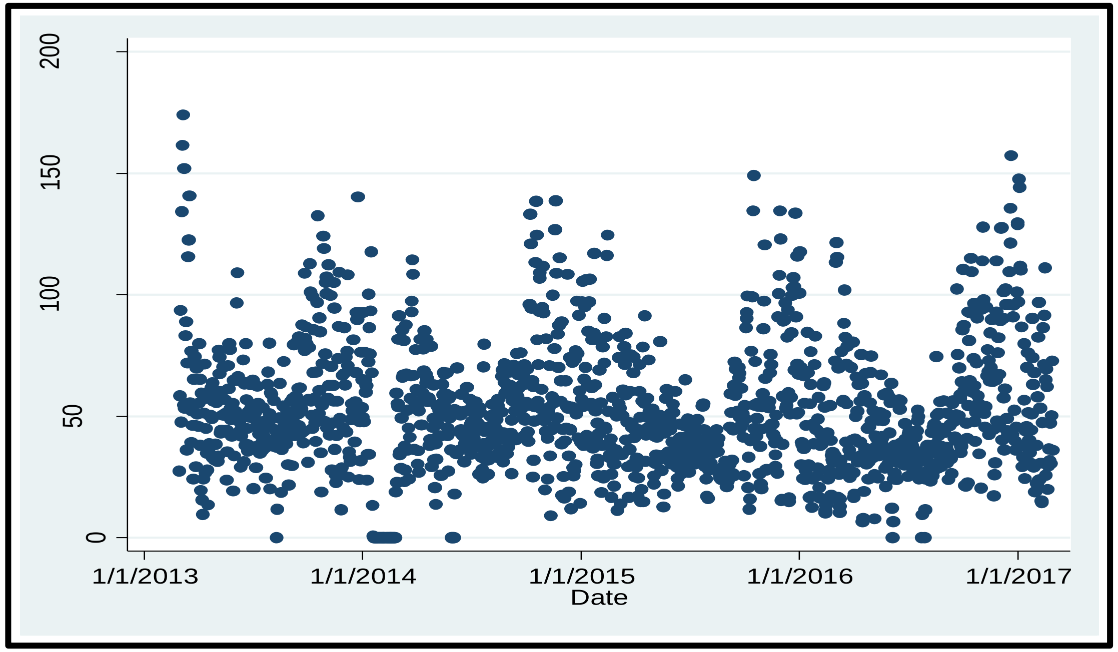


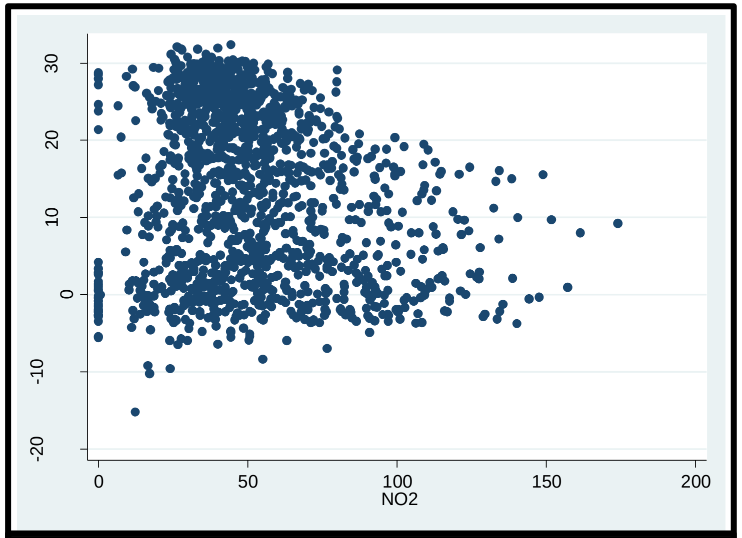


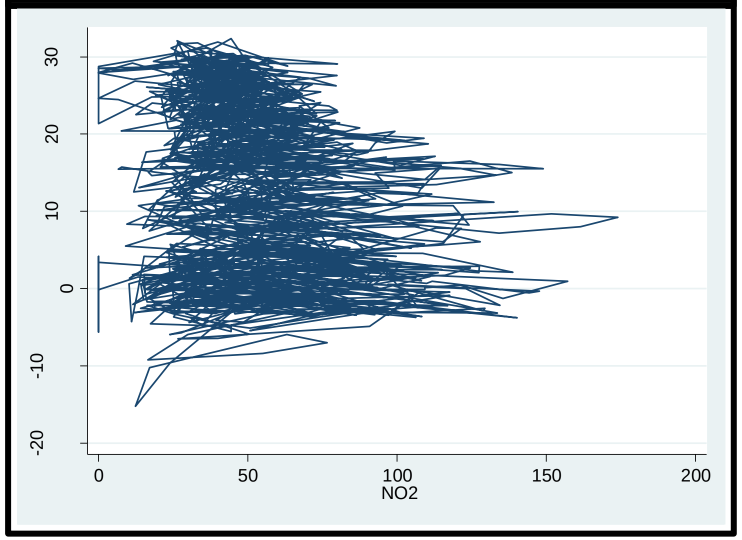


Temp vs NO2

Observation: The data doesn't show any significant change. Therefore there is no correlation between temperature and NO2.

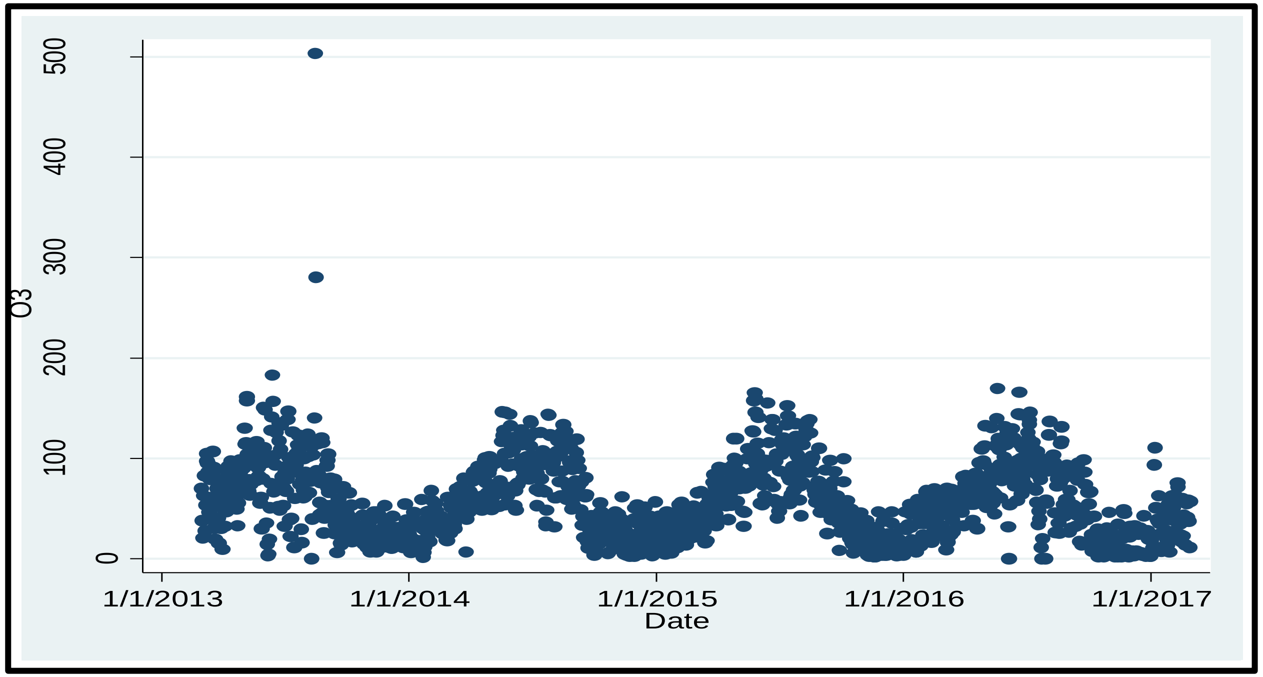


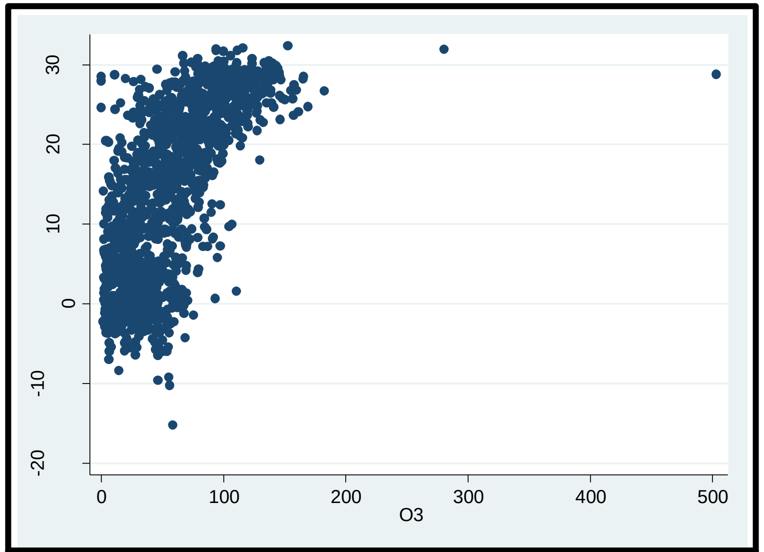


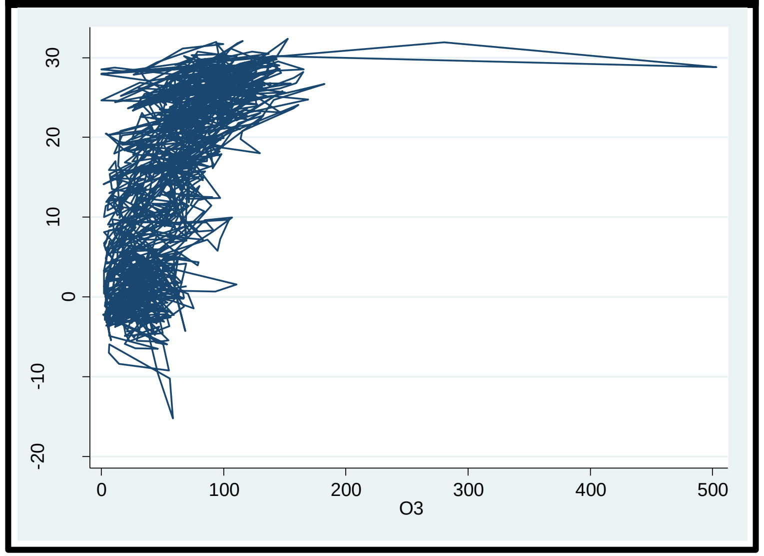


Temp vs O3

Observation: Similarly, for a section of the data, as O3 rises, temperature rises. This shows the relationship between temperature and O3 is highly positively correlated for a certain section.



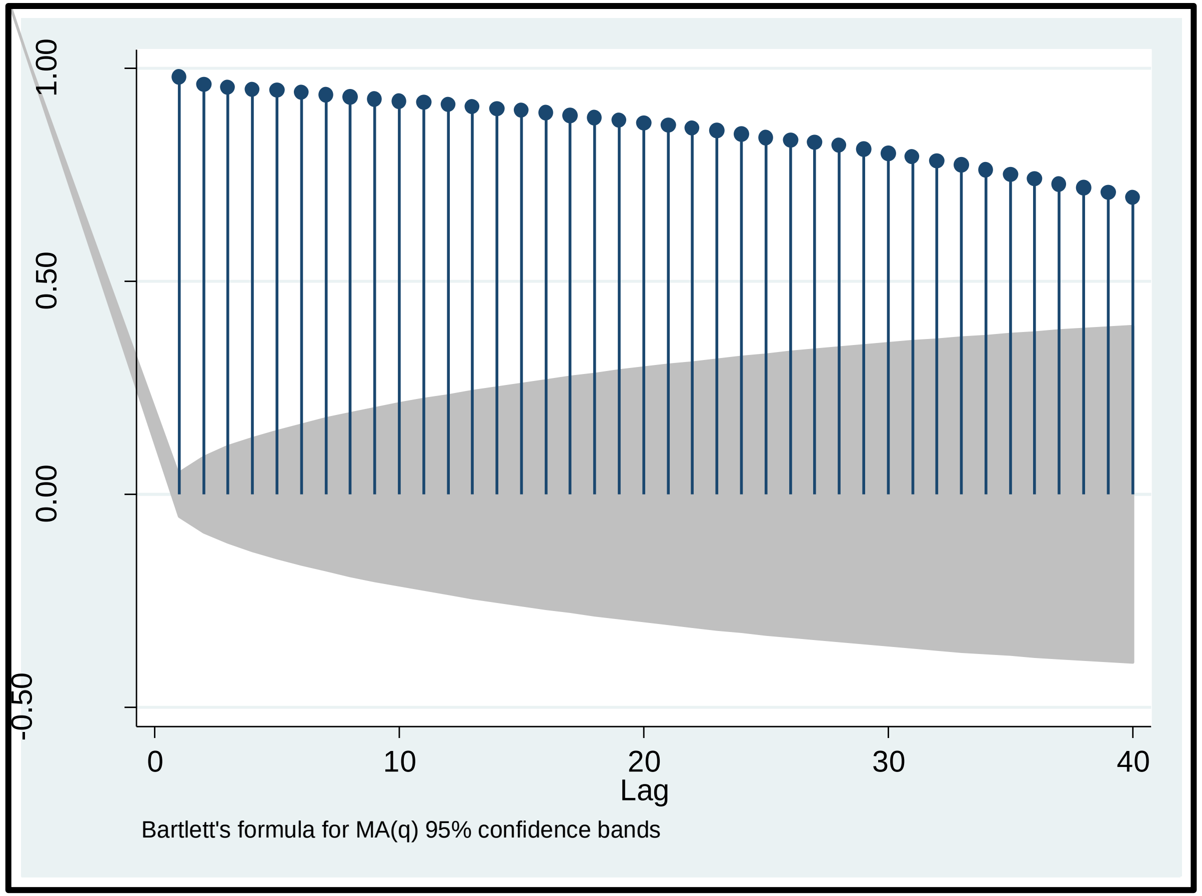




Auto Correlation

Observation: Temperature Variable is heavily Autocorrelated, since all the lag values have autocorrelations beyond the confidence level region of 95%.

As correlation is above grey area, it is statistically significant.



Auto Correlation Table

Observations:

- Correlation drops to lowest (0.832) at **12th lag**.
- Correlation again rises to highest point (0.9553) at **24th lag**.
- Correlation drops to lowest (0.8152) at **36th lag**.

34 . corrgram TEMP

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.9801	0.9801	1406.2	0.0000		
2	0.9625	0.0530	2763.4	0.0000		
3	0.9546	0.2407	4099.3	0.0000		
4	0.9504	0.1237	5424.4	0.0000		
5	0.9479	0.1382	6743.5	0.0000		
6	0.9444	0.0498	8053.6	0.0000		
7	0.9392	0.0242	9350.4	0.0000		
8	0.9336	0.0109	10633	0.0000		
9	0.9292	0.0399	11904	0.0000		

Friday September 9 23:09:46 2022 Page 3

LAG	AC	PAC	Q	Prob>Q
10	0.9236	-0.0341	13160	0.0000
11	0.9201	0.0619	14408	0.0000
12	0.9156	-0.0248	15645	0.0000
13	0.9105	0.0081	16868	0.0000
14	0.9056	-0.0011	18080	0.0000
15	0.9015	0.0218	19281	0.0000
16	0.8962	-0.0421	20469	0.0000
17	0.8907	-0.0063	21644	0.0000
18	0.8842	-0.0514	22802	0.0000
19	0.8775	-0.0094	23943	0.0000
20	0.8715	-0.0239	25070	0.0000
21	0.8658	-0.0026	26182	0.0000
22	0.8601	-0.0150	27281	0.0000
23	0.8540	-0.0061	28365	0.0000
24	0.8465	-0.0572	29431	0.0000
25	0.8384	-0.0323	30477	0.0000
26	0.8319	-0.0050	31508	0.0000
27	0.8260	-0.0087	32525	0.0000
28	0.8191	-0.0306	33526	0.0000
29	0.8104	-0.0533	34506	0.0000
30	0.8008	-0.0545	35464	0.0000
31	0.7919	-0.0229	36401	0.0000
32	0.7832	-0.0482	37319	0.0000
33	0.7737	-0.0514	38215	0.0000
34	0.7622	-0.1005	39085	0.0000
35	0.7513	-0.0259	39931	0.0000
36	0.7399	-0.0827	40752	0.0000
37	0.7293	-0.0166	41551	0.0000
38	0.7196	-0.0257	42329	0.0000
39	0.7087	-0.0424	43084	0.0000
40	0.6966	-0.0580	43813	0.0000

Simple Regression model

$Temp_t = \alpha + \beta_1 PM2.5_t + \beta_2 PM10_t + \beta_3 SO_{2t} + \beta_4 NO_{2t} + \beta_5 CO_t + \beta_6 O_{3t} + \mu$

Observation : On seeing the P-value of all the independent variables, we find that all the variables are significant at 5% level of significance.

R-squared is 0.565

```
. regress TEMP PM25 PM10 SO2 NO2 CO O3
```

Source	SS	df	MS
Model	98041.6871	6	16340.2812
Residual	75504.4924	1,454	51.9288118
Total	173546.179	1,460	118.867246

Number of obs	=	1,461
F(6, 1454)	=	314.67
Prob > F	=	0.0000
R-squared	=	0.5649
Adj R-squared	=	0.5631
Root MSE	=	7.2062

TEMP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PM25	.0311779	.0071316	4.37	0.000	.0171887	.0451672
PM10	-.0132851	.0061558	-2.16	0.031	-.0253603	-.0012098
SO2	-.1825716	.0122183	-14.94	0.000	-.2065389	-.1586044
NO2	.0805904	.0103486	7.79	0.000	.0602905	.1008902
CO	-.0017569	.0003306	-5.31	0.000	-.0024054	-.0011084
O3	.176298	.0055322	31.87	0.000	.1654461	.1871499
_cons	3.894377	.6183003	6.30	0.000	2.681522	5.107233

Detecting autocorrelation using Durbin Watson, Q-Statistic, Breusch-Godfrey LM test

Observation: Highly positively correlated

P-Value is 0 therefore rejecting H_0

```
. dwstat

Durbin-Watson d-statistic(  7,  1461) =  .5203704
```

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	845.496	1	0.0000

H0: no serial correlation

```
. wntestq res

Portmanteau test for white noise

 Portmanteau (Q) statistic = 11739.4846
 Prob > chi2(40)           =     0.0000
```

Autoregressive Model

$Temp_t = \alpha + \beta_1 PM2.5_t + \beta_2 PM10_t + \beta_3 SO_{2t} + \beta_4 NO_{2t} + \beta_5 CO_t + \beta_6 O_{3t} + \beta_7 Temp_{t-1} + \mu$

R squared increased from 0.56 to 0.96

Coefficient of lagged Temp is 0.9

On seeing the P- value of all the independent variables, we find that except PM10, all the variables are significant at 5% level of significance.

```
. regress TEMP PM25 PM10 SO2 NO2 CO O3 L.TEMP
```

Source	SS	df	MS
Model	167234.207	7	23890.601
Residual	6159.62522	1,452	4.24216613
Total	173393.832	1,459	118.844299

Number of obs	=	1,460
F(7, 1452)	=	5631.70
Prob > F	=	0.0000
R-squared	=	0.9645
Adj R-squared	=	0.9643
Root MSE	=	2.0597

TEMP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PM25	.0048367	.0020488	2.36	0.018	.0008178	.0088556
PM10	.0001048	.0017629	0.06	0.953	-.0033534	.0035629
SO2	-.0095908	.0037457	-2.56	0.011	-.0169382	-.0022433
NO2	.011491	.0030069	3.82	0.000	.0055926	.0173894
CO	-.0002846	.0000952	-2.99	0.003	-.0004713	-.0000978
O3	.0184477	.0020077	9.19	0.000	.0145094	.0223859
TEMP						
L1.	.9319525	.0073011	127.65	0.000	.9176308	.9462743
_cons	-.5888481	.1802433	-3.27	0.001	-.9424132	-.2352831

Detecting autocorrelation using Durbin Watson, Q-Statistic, Breusch-Godfrey LM test

Observations: Highly positively correlated.

P-Value is 0 therefore rejecting H_0 .

```
. estat durbinalt

Durbin's alternative test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	1.851	1	0.1737

H0: no serial correlation

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	1.860	1	0.1726

H0: no serial correlation

```
. wntestq res

Portmanteau test for white noise

 Portmanteau (Q) statistic = 11739.4846
 Prob > chi2(40)           =     0.0000
```

Prais –Winsten and Cochrane –Orcutt regression model

Observations: Since p-values of NO2 is greater than 0.05, it is not significant.

On seeing transformed Durbin Watson Statistic, we can conclude that there is no autocorrelation.

```
. prais TEMP PM25 PM10 SO2 NO2 O3, corc

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.7392
Iteration 2:  rho = 0.9722
Iteration 3:  rho = 0.9798
Iteration 4:  rho = 0.9798
Iteration 5:  rho = 0.9798

Cochrane-Orcutt AR(1) regression -- iterated estimates

      Source |       SS           df       MS      Number of obs   =     1,460
-------------+----------------------------------   F(5, 1454)      =     36.74
       Model |  750.268361         5  150.053672   Prob > F        =    0.0000
    Residual |  5937.64238     1,454  4.08366051   R-squared       =    0.1122
-------------+----------------------------------   Adj R-squared   =    0.1091
       Total |  6687.91074     1,459  4.58390044   Root MSE        =    2.0208

------------------------------------------------------------------------------
        TEMP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PM25 |  -.0050339   .0021359    -2.36   0.019    -.0092237    -.000844
        PM10 |   .0072812   .0018916     3.85   0.000     .0035707    .0109917
         SO2 |   -.012828     .00444    -2.89   0.004    -.0215375   -.0041185
         NO2 |   .0047643   .0037435     1.27   0.203     -.002579    .0121075
          O3 |   .0257456   .0021443    12.01   0.000     .0215394    .0299518
       _cons |   12.17396   2.625702     4.64   0.000     7.023394    17.32453
-------------+----------------------------------------------------------------
         rho |   .9797976
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.518768
Durbin-Watson statistic (transformed) 2.135747
```

Solution to Problem

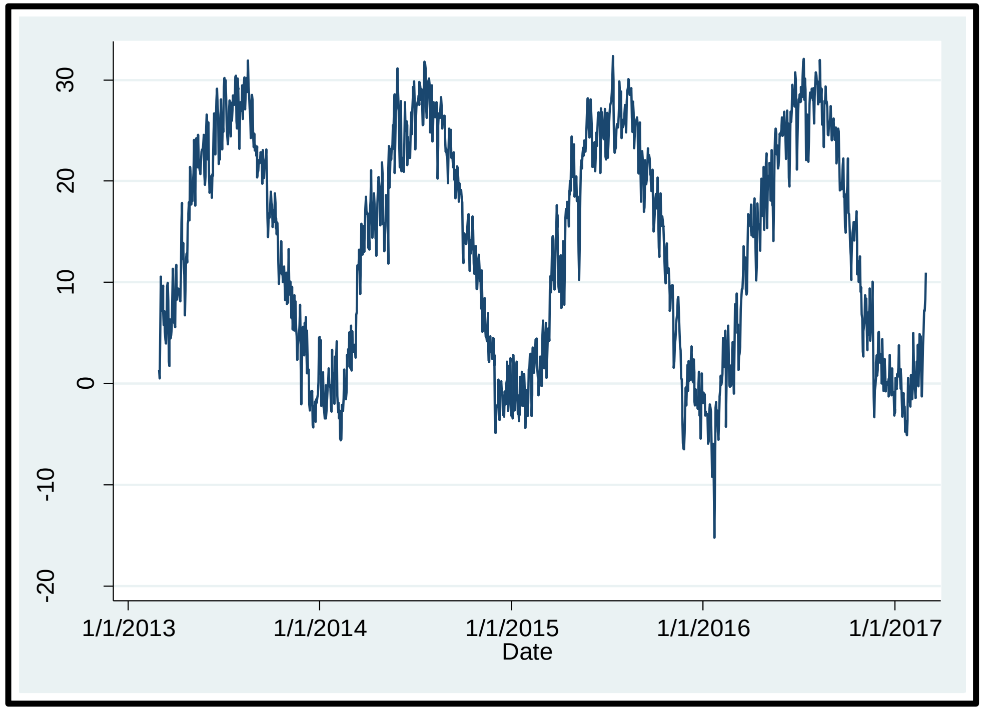
To solve this problem, in the places where values were not available we have inserted the average of all the values of the variable so that mean does not gets disturbed.

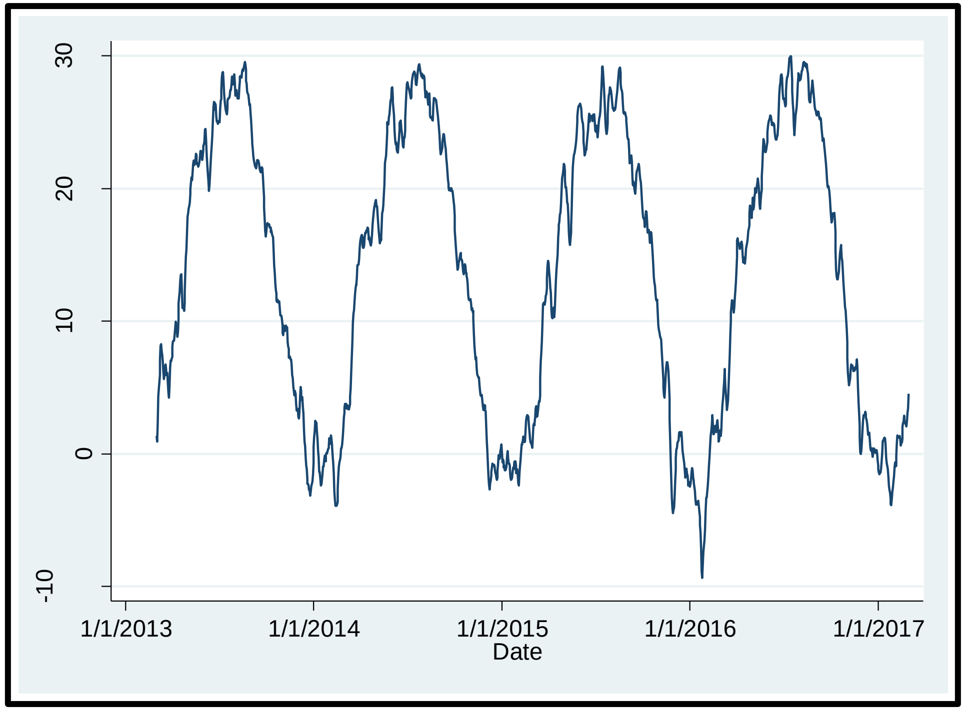
No	year	month	day	hour	PM2.5	PM10	SO2	NO2	CO	O3	TEMP	PRES	DEWP	RAIN
1	2013	3	1	0	9	9	3	17	300	89	-0.5	1024.5	-21.4	0
2	2013	3	1	1	4	4	3	16	300	88	-0.7	1025.1	-22.1	0
3	2013	3	1	2	7	7		17	300	60	-1.2	1025.3	-24.6	0
4	2013	3	1	3	3	3	5	18			-1.4	1026.2	-25.5	0
5	2013	3	1	4	3	3	7		200	84	-1.9	1027.1	-24.5	0
6	2013	3	1	5	4	4	9	25	300	78	-2.4	1027.5	-21.3	0
7	2013	3	1	6	5	5	10	29	400	67	-2.5	1028.2	-20.4	0
8	2013	3	1	7	3	6	12	40	400	52	-1.4	1029.5	-20.4	0
9	2013	3	1	8	3	6	12	41	500	54	-0.3	1030.4	-21.2	0
10	2013	3	1	9	3	6	9	31	400	69	0.4	1030.5	-23.3	0
11	2013	3	1	10	3	6	7	19	300	82	1.4	1030.2	-22.5	0
12	2013	3	1	11	6	6	7	19	400	83	2.9	1029.8	-22.9	0
13	2013	3	1	12	3	6	6	18	300	86	4	1028.6	-21.2	0
14	2013	3	1	13	3	6	5	16	300	91	5	1027.8	-21.2	0
15	2013	3	1	14	3	6	5	16	300	92	6.2	1027.6	-22.2	0
16	2013	3	1	15	3	10	5	17	300	92	6	1027.7	-21.3	0
17	2013	3	1	16	9	17	6	19	300	91	5.6	1027.7	-20.7	0
18	2013	3	1	17	9	22	9	21	400	87	4.4	1028.2	-20.9	0
19	2013	3	1	18	11	23	8	28	400	79	3.2	1029.4	-20.3	0
20	2013	3	1	19	13	17	12	42	600	63	3	1030.1	-19.7	0

PRSA_Data_Dongsi_20130301-20170

dy Accessibility: Unavailable Average: 18.53110661 Count: 34402 Sum: 637488.5984

7 Day Moving Average



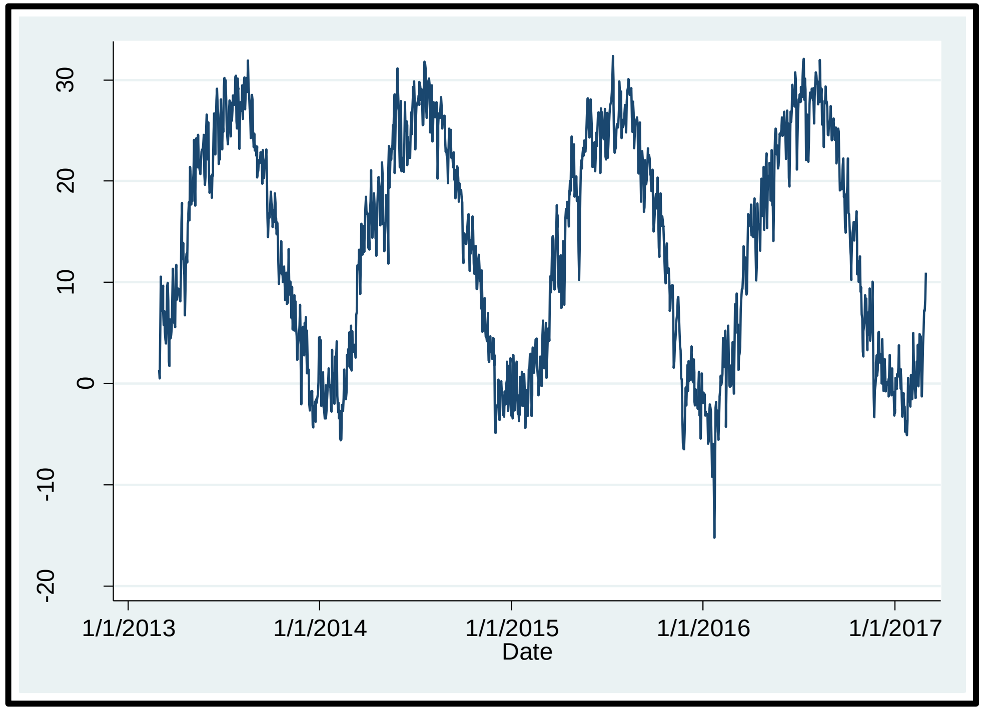


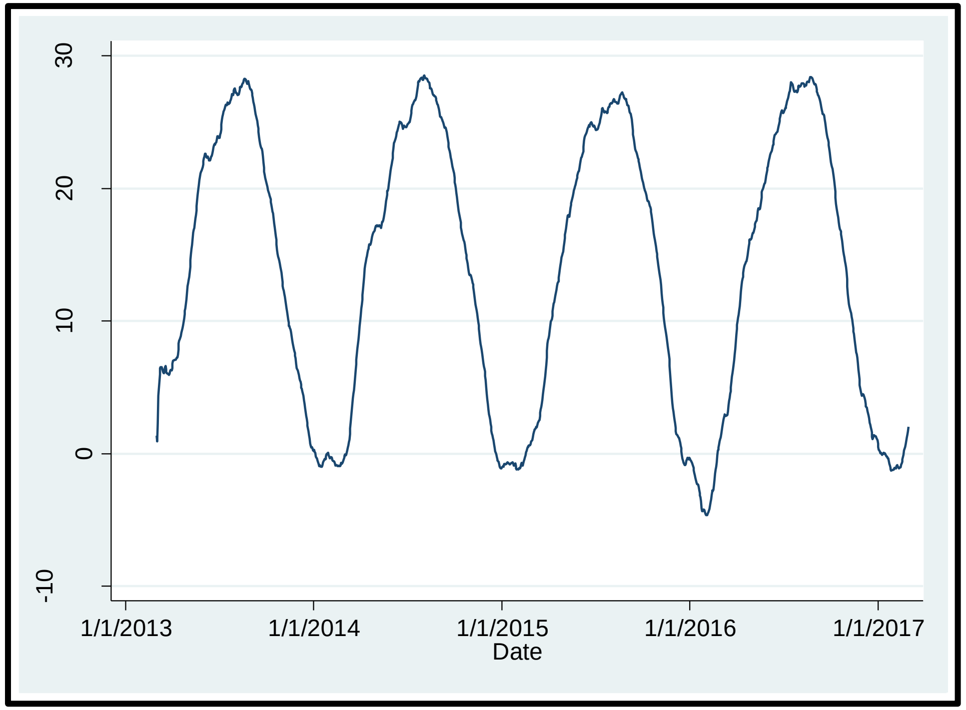
The moving average for time period is the arithmetic mean of the most recent observations where equal weights are assigned to each observation.

The objective of doing moving average is to smooth out the temperature date over time.

In this graph, we can see the seasonal variation with time in 7 day lag.

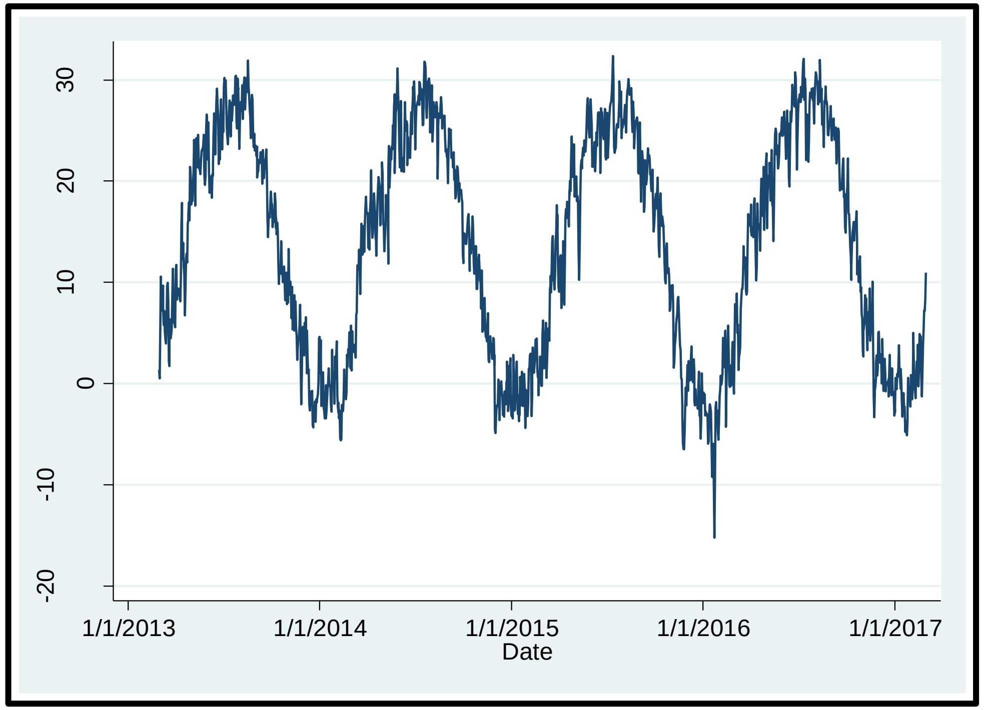
30 Day Moving Average

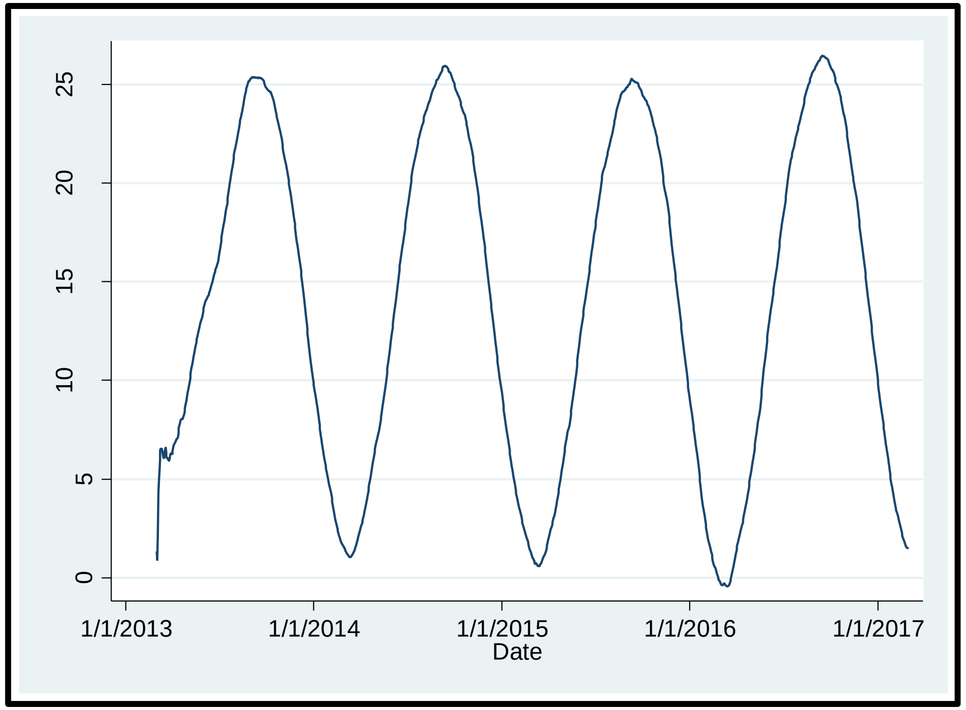




The longer the period for the moving average, the greater the lag.
Different periods are chosen to calculate moving average based on the objective.
Short-term moving average used for short-term analysis. This graph again shows seasonality and a similar trend as observed before.

120 Day Moving Average





Here we have done 120 day moving average for long-term analysis.

Over the long-term also, seasonality in our data can be observed.

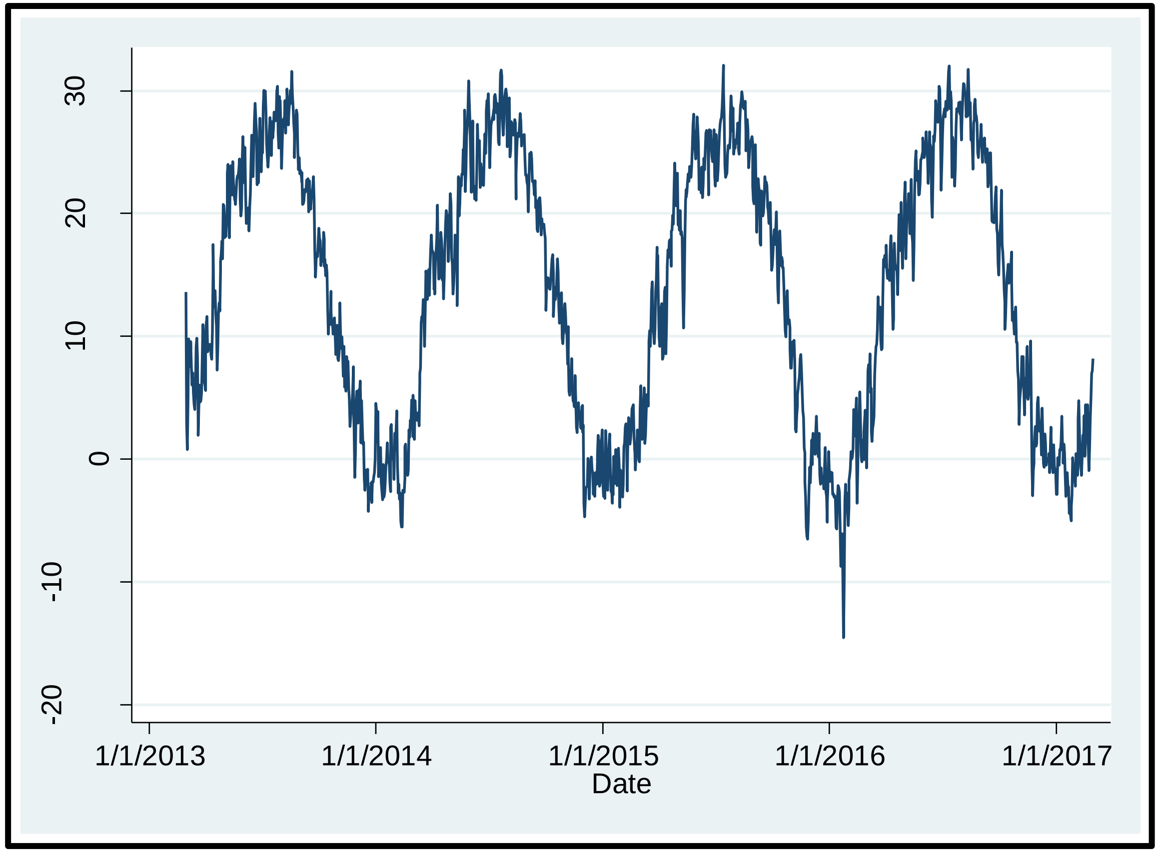
Through this we can see the impacts of random and short-term fluctuations of the variables on the temperature over a specific period of time.

Single Exponential Smooth

Exponential smoothening puts more emphasis on the recent data points. It uses weighted average calculations.

Observation: The value optimal exponential coefficient is 0.87 which shows that it tracks the data closely by giving more weight to the recent data.

```
. tssmooth exponential ewma = TEMP

computing optimal exponential  coefficient (0,1)

optimal exponential coefficient =        0.8718
sum-of-squared residuals        =     6856.8045
root mean squared error         =     2.1663857
```

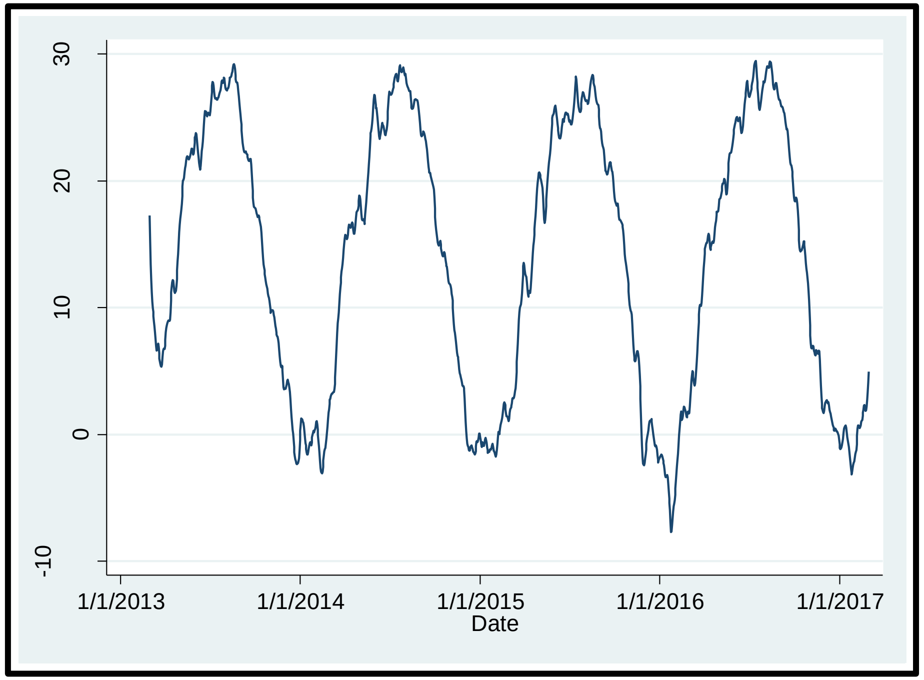


Double Exponential Smoothing

This graph shows Temp after double exponential smoothing. It is used for forecasting the time series when the data has a linear trend and no seasonal pattern. It is a special case

of holt's exponential smooth where alpha=beta

Observation: The value optimal exponential coefficient is 0.27 which shows that it tracks the data closely by giving less weight to the recent data.

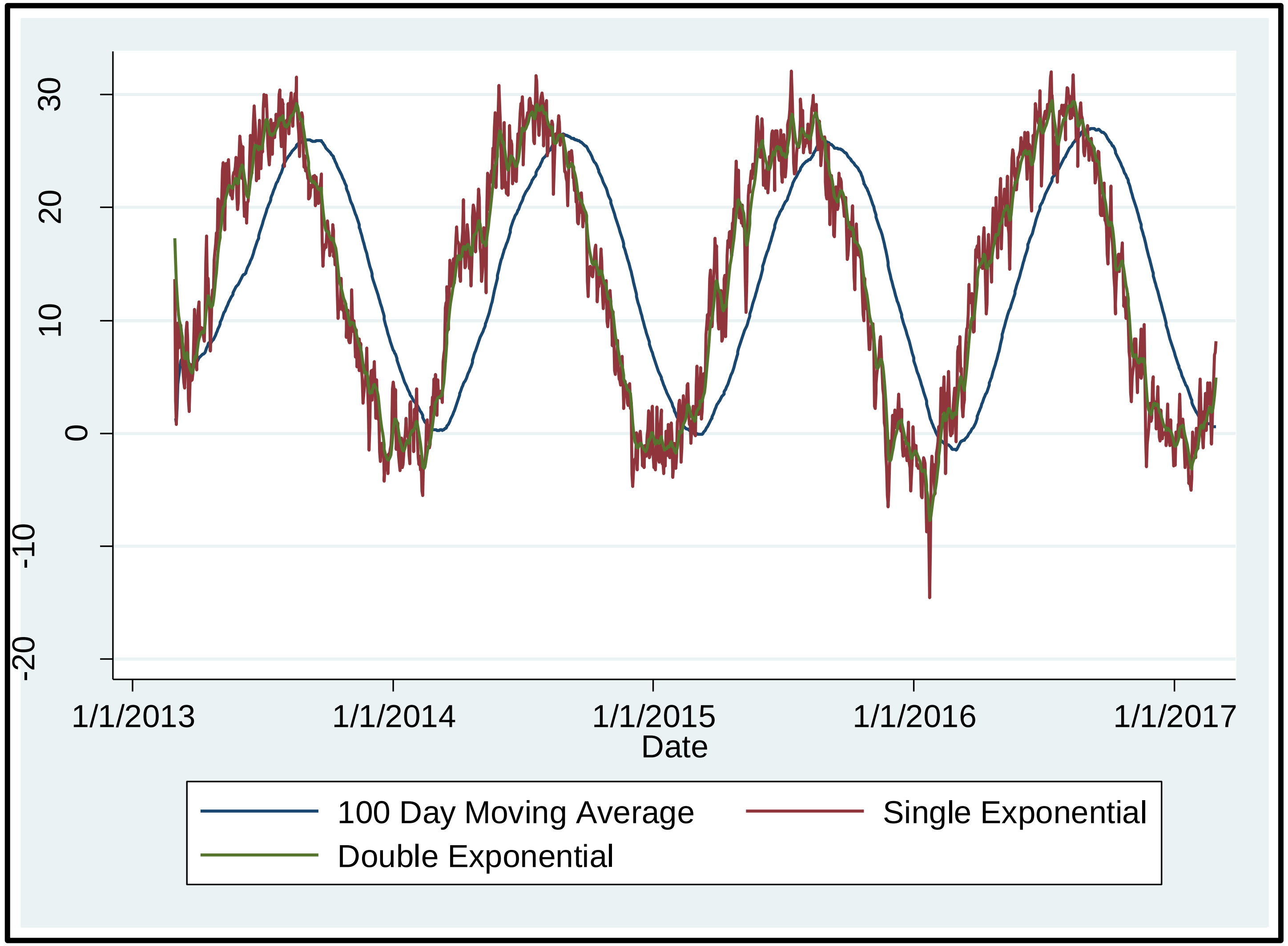



```
. tssmooth dexponential dewma = TEMP
computing optimal double-exponential coefficient (0,1)

optimal double-exponential coefficient =       0.2791
sum-of-squared residuals               =    9390.7927
root mean squared error                =    2.5352805
```

Comparison Graph

Observation: The graph shows comparison of 100 day moving average, single exponential and double exponential.

Blue solid line shows 100 day moving average, red line shows single exponential and green line shows double exponential.



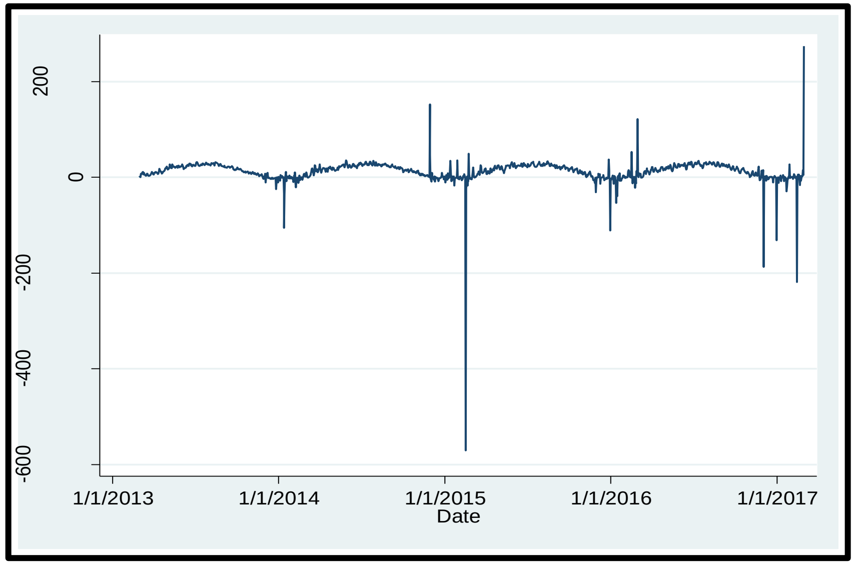
Holt Winter (without seasonal component) smoothing optimal parameter

```
. tssmooth hwinters hw = TEMP
computing optimal weights

Iteration 0:    penalized RSS = -10387.423
Iteration 1:    penalized RSS = -7446.7538
Iteration 2:    penalized RSS =  -7009.369
Iteration 3:    penalized RSS = -7006.6979
Iteration 4:    penalized RSS = -7006.1287
Iteration 5:    penalized RSS = -7005.2612
Iteration 6:    penalized RSS = -7005.2607
Iteration 7:    penalized RSS = -7005.2607  (backed up)
Iteration 8:    penalized RSS = -7005.2607  (backed up)
Iteration 9:    penalized RSS = -7005.2607  (backed up)
Iteration 10:   penalized RSS = -7005.2607  (backed up)
Iteration 11:   penalized RSS = -7005.2607  (backed up)
Iteration 12:   penalized RSS = -7005.2607  (backed up)
Iteration 13:   penalized RSS = -7005.2607  (backed up)
Iteration 14:   penalized RSS = -7005.2607  (backed up)
Iteration 15:   penalized RSS = -7005.2607  (backed up)
Iteration 16:   penalized RSS = -7005.2607  (backed up)
Iteration 17:   penalized RSS = -7005.2607  (backed up)
Iteration 18:   penalized RSS = -7005.2607  (backed up)
Iteration 19:   penalized RSS = -7005.2607  (backed up)
Iteration 20:   penalized RSS = -7005.2607  (backed up)
Iteration 21:   penalized RSS = -7005.2607  (backed up)
Iteration 22:   penalized RSS = -7005.2607  (backed up)
Iteration 23:   penalized RSS = -7005.2607  (backed up)
Iteration 24:   penalized RSS = -7005.2607  (backed up)
Iteration 25:   penalized RSS = -7005.2607  (backed up)
Iteration 26:   penalized RSS = -7005.2607  (backed up)
```

Holt-Winters (with seasonal component) with optimal parameters

Holt Winters smoothing technique is used to handle the problem of seasonality.

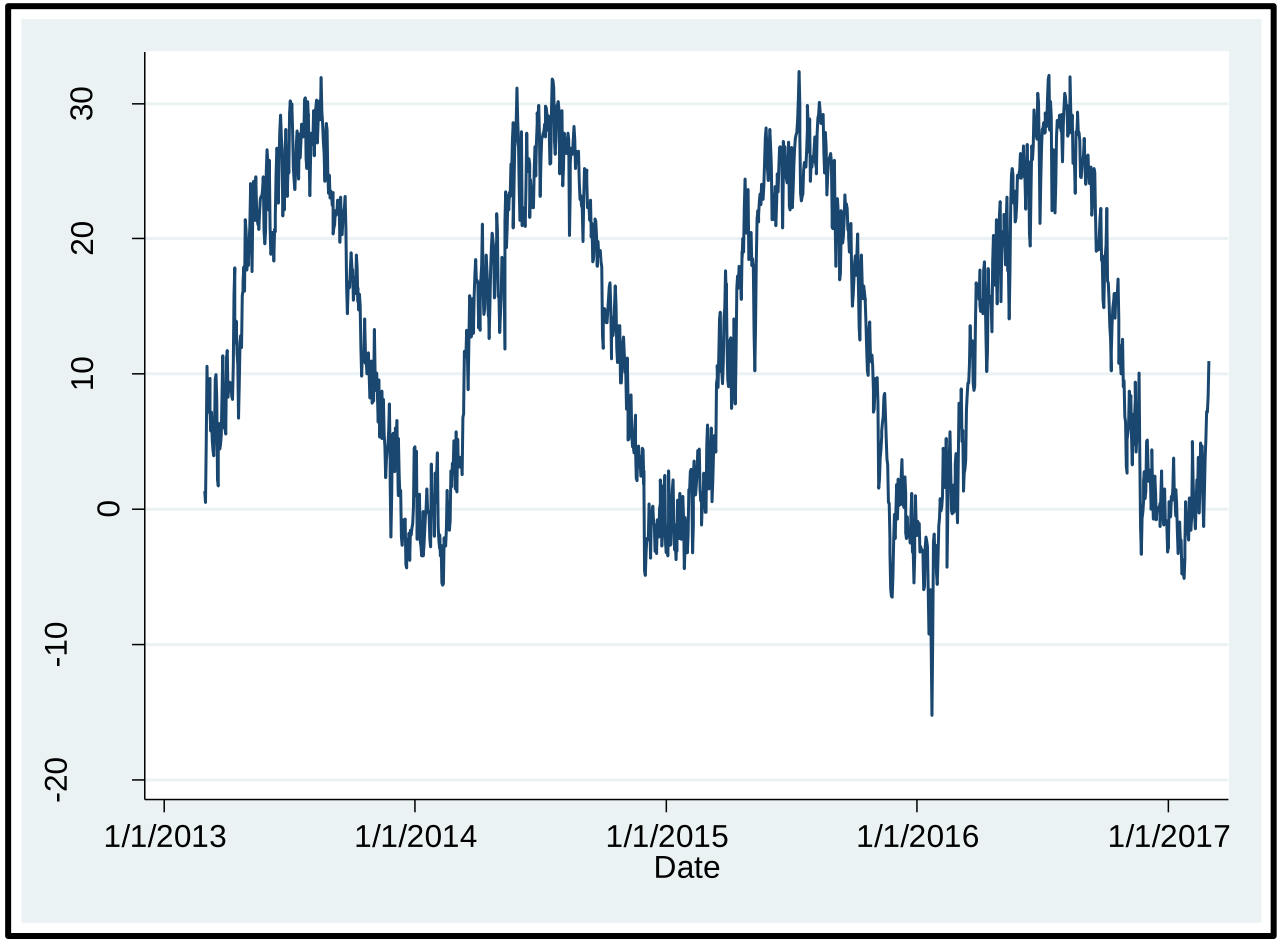
```
. tssmooth shwinters shw = TEMP
computing optimal weights

Iteration 0:   penalized RSS = -1010720.4  (not concave)
Iteration 1:   penalized RSS = -887140.03
Iteration 2:   penalized RSS = -640577.62
Iteration 3:   penalized RSS = -613598.14
Iteration 4:   penalized RSS = -605571.59
Iteration 5:   penalized RSS = -604615.51
Iteration 6:   penalized RSS = -604596.24
Iteration 7:   penalized RSS = -604596.19

Optimal weights:
                              alpha = 0.5023
                               beta = 0.4970
                              gamma = 0.5004
penalized sum-of-squared residuals = 604596.2
          sum-of-squared residuals = 604596.2
          root mean squared error  = 20.34265
```



Stationarity

Conditions For Stationarity :

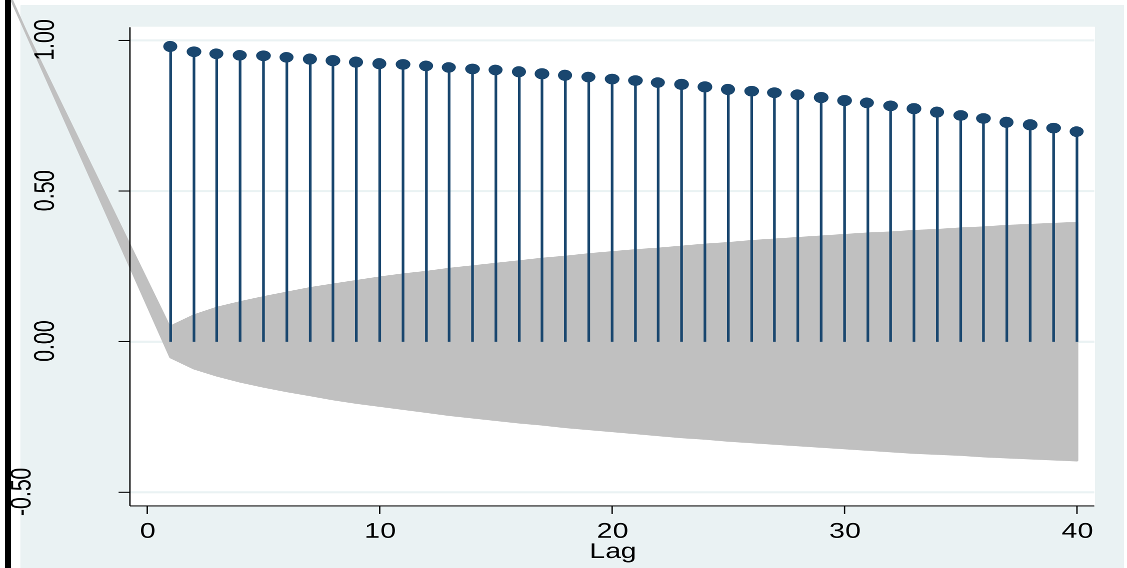
- Constant Mean
- Constant Variance
- No Seasonality



Our data set has constant mean, constant variance but there is seasonality in our data hence our data is not stationary.

Auto Correlation and Correlogram

For stationary time series, values degrade to zero quickly whereas in non-stationary, values degrade more slowly. For our data, Autocorrelation plot indicates strong persistence and slow degradation, so this implies that series is not stationary. But we can't rely entirely on correlogram



Bartlett's formula for MA(q) 95% confidence bands

```
. corrgram temp_n
(note: time series has 8 gaps)
```

LAG	AC	PAC	Q	Prob>Q
1	0.9938	0.9942	34614	0.0000
2	0.9819	-0.5239	68401	0.0000
3	0.9657	-0.1372	1.0e+05	0.0000
4	0.9467	-0.0383	1.3e+05	0.0000
5	0.9261	0.0410	1.6e+05	0.0000
6	0.9054	0.0711	1.9e+05	0.0000
7	0.8857	0.1013	2.2e+05	0.0000
8	0.8681	0.1084	2.5e+05	0.0000
9	0.8534	0.1148	2.7e+05	0.0000
10	0.8424	0.1176	3.0e+05	0.0000
11	0.8352	0.0940	3.2e+05	0.0000
12	0.8320	0.1032	3.4e+05	0.0000
13	0.8327	0.1066	3.7e+05	0.0000
14	0.8374	0.1255	3.9e+05	0.0000
15	0.8456	0.1371	4.2e+05	0.0000
16	0.8571	0.1591	4.4e+05	0.0000
17	0.8713	0.1430	4.7e+05	0.0000
18	0.8871	0.1138	5.0e+05	0.0000
19	0.9036	0.1071	5.3e+05	0.0000
20	0.9197	0.0880	5.6e+05	0.0000
21	0.9341	0.0519	5.9e+05	0.0000
22	0.9457	0.0132	6.2e+05	0.0000
23	0.9531	-0.0534	6.5e+05	0.0000
24	0.9553	-0.1143	6.8e+05	0.0000
25	0.9516	-0.1628	7.1e+05	0.0000
26	0.9427	-0.0460	7.5e+05	0.0000
27	0.9297	-0.0043	7.8e+05	0.0000
28	0.9139	0.0180	8.0e+05	0.0000
29	0.8965	0.0113	8.3e+05	0.0000
30	0.8787	0.0034	8.6e+05	0.0000
31	0.8615	0.0026	8.9e+05	0.0000
32	0.8461	0.0079	9.1e+05	0.0000
33	0.8333	0.0109	9.4e+05	0.0000
34	0.8237	0.0113	9.6e+05	0.0000
35	0.8176	0.0127	9.8e+05	0.0000
36	0.8152	0.0049	1.0e+06	0.0000
37	0.8164	0.0082	1.0e+06	0.0000
38	0.8212	0.0272	1.1e+06	0.0000
39	0.8294	0.0361	1.1e+06	0.0000
40	0.8407	0.0517	1.1e+06	0.0000

Dickey Fuller Test for Stationarity (Unit root Differences)

Critical Value for our test statistic is more than the critical values at 99% , 95% and 90% Significance level.

Also P- value is less than any significance level. We **reject** the null hypothesis.

Thus, the series exhibits Stationarity. This is because our data is strongly auto-correlated. Hence we'll use augmented-dickey fuller test which removes auto-correlation from the series.

$$H_0 = Unit\ root\ is\ present$$
$$H_a = No\ Unit\ root\ (Stationarity)$$

```
. dfuller TEMP

Dickey-Fuller test for unit root                   Number of obs   =      1460

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.868            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0023
```

Dickey Fuller Test for Stationarity (3 lagged differences)

Critical Value for our test statistic is less than the critical values at 99% , 95% and 90% Significance level.

Also P- value is more than any significance level. We **fail to reject** the null hypothesis.

Thus, the series does not exhibits Stationarity.

```
. dfuller TEMP, lags(3) trend regress
```

Augmented Dickey-Fuller test for unit root   Number of obs = 1457

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.515	-3.960	-3.410	-3.120

MacKinnon approximate p-value for Z(t) = 0.3207

D.TEMP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
TEMP						
L1.	-.0126761	.0050411	-2.51	0.012	-.0225646	-.0027875
LD.	-.1052628	.0261065	-4.03	0.000	-.1564734	-.0540522
L2D.	-.2508423	.0252814	-9.92	0.000	-.3004343	-.2012503
L3D.	-.1237733	.0259521	-4.77	0.000	-.174681	-.0728657
_trend	-.0000841	.0001292	-0.65	0.515	-.0003375	.0001694
_cons	.2372249	.1352288	1.75	0.080	-.02804	.5024898

Augmented Dickey Fuller Test for Stationarity

Critical Value for our test statistic is less than the critical values at 99% , 95% and 90% Significance level.

Also P- value is more than any significance level. We **fail to reject** the null hypothesis.

Thus, the series does not exhibits Stationarity.

$$H_0 = Unit\ root\ is\ present, \boldsymbol{alpha = 1}$$
$$H_a = No\ Unit\ root\ (Stationarity)$$

```
. dfuller TEMP, lags(3) trend

Augmented Dickey-Fuller test for unit root         Number of obs   =      1457

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.515            -3.960            -3.410            -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3207
```

Augmented Dickey Fuller Test for Stationarity

Critical Value for our test statistic is less than the critical values at 99% , 95% and 90% Significance level.

Also P- value is more than any significance level. We **fail to reject** the null hypothesis.

Thus, the series does not exhibits Stationarity.

$$H_0 = Unit\ root\ is\ present$$
$$H_a = No\ Unit\ root\ (Stationarity)$$

```
. dfuller TEMP, lags(3) trend regress

Augmented Dickey-Fuller test for unit root         Number of obs   =      1457
```

		Interpolated Dickey-Fuller		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.515	-3.960	-3.410	-3.120

MacKinnon approximate p-value for Z(t) = 0.3207

D.TEMP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
TEMP						
L1.	-.0126761	.0050411	-2.51	0.012	-.0225646	-.0027875
LD.	-.1052628	.0261065	-4.03	0.000	-.1564734	-.0540522
L2D.	-.2508423	.0252814	-9.92	0.000	-.3004343	-.2012503
L3D.	-.1237733	.0259521	-4.77	0.000	-.174681	-.0728657
_trend	-.0000841	.0001292	-0.65	0.515	-.0003375	.0001694
_cons	.2372249	.1352288	1.75	0.080	-.02804	.5024898

Dickey-Fuller Generalized least square test for unit root

$$H_0 = \text{Unit root is present}$$
$$H_a = \text{No Unit root (Stationarity)}$$

Observations: From the data we can see that the null hypothesis at 5% level cannot be rejected, not even at 10% significance level from Lag 2 Onwards.

For Lag 1 , Null can be rejected at 5% and 10% & For Lag 2 , Null can be rejected at 10%.

```
. dfgls TEMP

DF-GLS for TEMP                                          Number of obs =   1437
Maxlag = 23 chosen by Schwert criterion
```

[lags]	DF-GLS tau Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
23	-1.761	-3.480	-2.829	-2.543
22	-1.656	-3.480	-2.830	-2.544
21	-1.649	-3.480	-2.831	-2.545
20	-1.625	-3.480	-2.832	-2.546
19	-1.623	-3.480	-2.834	-2.547
18	-1.580	-3.480	-2.835	-2.548
17	-1.571	-3.480	-2.836	-2.549
16	-1.481	-3.480	-2.837	-2.550
15	-1.463	-3.480	-2.838	-2.551
14	-1.388	-3.480	-2.839	-2.552
13	-1.436	-3.480	-2.840	-2.553
12	-1.443	-3.480	-2.841	-2.554
11	-1.463	-3.480	-2.842	-2.555
10	-1.423	-3.480	-2.843	-2.556
9	-1.532	-3.480	-2.844	-2.556

123 Saturday October 8 21:44:46 2022 Page 4

[lags]	DF-GLS tau Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
8	-1.475	-3.480	-2.845	-2.557
7	-1.543	-3.480	-2.846	-2.558
6	-1.564	-3.480	-2.847	-2.559
5	-1.603	-3.480	-2.848	-2.560
4	-1.697	-3.480	-2.848	-2.561
3	-1.979	-3.480	-2.849	-2.562
2	-2.256	-3.480	-2.850	-2.562
1	-2.910	-3.480	-2.851	-2.563

```
Opt Lag (Ng-Perron seq t) = 23 with RMSE  2.013386
Min SC   =  1.445134 at lag  4 with RMSE  2.033826
Min MAIC =  1.427068 at lag 10 with RMSE  2.024065
```

Phillips-Perron Test for unit root

$$H_0 = Unit\ root\ is\ present$$
$$H_a = No\ Unit\ root\ (Stationarity)$$

As we can observe from the data, P- value is 0. So, we fail to reject the null hypothesis and conclude that there is unit root and that the series in not stationary.

```
. pperron TEMP, regress

Phillips-Perron test for unit root                 Number of obs   =      1460
                                                   Newey-West lags =         7

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(rho)          -14.685           -20.700           -14.100           -11.300
 Z(t)             -2.793            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0593

------------------------------------------------------------------------------
        TEMP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        TEMP |
         L1. |   .9801153   .0051412   190.64   0.000     .9700303    .9902003
             |
       _cons |   .2783334   .0898775     3.10   0.002     .1020304    .4546363
```

KPSS

$$H_0 = Stationary$$
$$H_a = Non - \ Stationary$$
$$(Difference\ stationary)$$

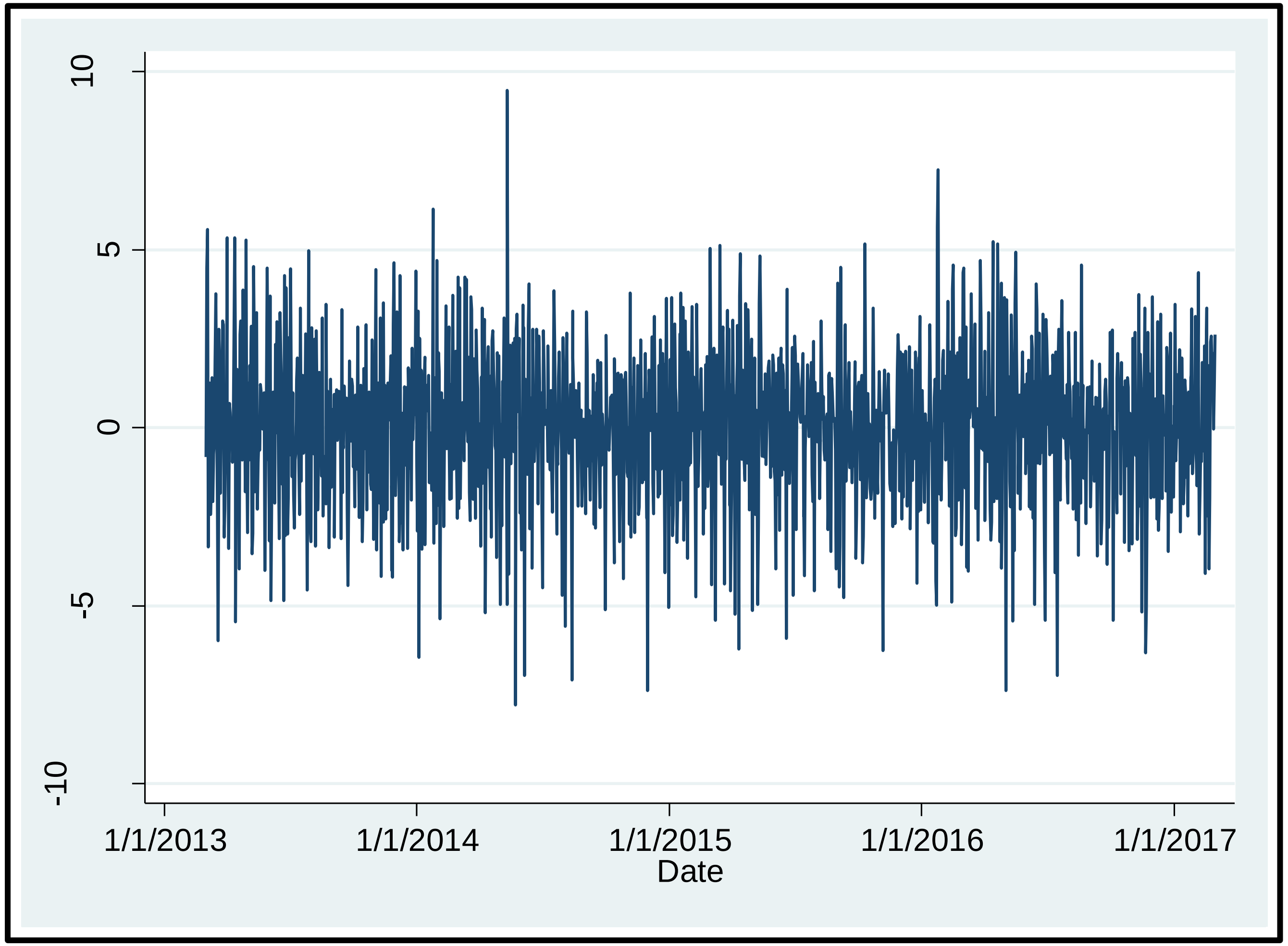
The KPSS test is based on linear regression. It breaks up a series into three parts: a deterministic trend (βt), a random walk (rt), and a stationary error (εt).

Since, the Test statistic is more than the Critical value at 1% at all lags, we reject the null hypothesis of Stationarity.

```
. kpss TEMP

KPSS test for TEMP

Maxlag = 7 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: TEMP is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0            2.48
    1            1.25
    2             .84
    3            .634
    4            .509
    5            .426
    6            .366
    7            .321
```

Correction of non-stationarity

After differencing, we have created a new variable Temp 1 and run the test.
From the graph, we can see that the data exhibits stationarity now as there is constant mean, constant variance, and no seasonality.
All the three conditions are satisfied



Dickey Fuller Test for Stationarity

Observations: With the trend term, we can see a slight difference in our results but it doesn't signify a trend. P- value is 0.0000 , therefore we can reject the null hypothesis and stationarity has been resolved.

```
. dfuller temp1, lags(3) trend regress

Augmented Dickey-Fuller test for unit root         Number of obs   =      1456

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)           -26.998            -3.960            -3.410            -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D.temp1      |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       temp1 |
         L1. |  -1.708098    .063267   -27.00   0.000    -1.832202   -1.583993
         LD. |   .5781724   .0514737    11.23   0.000     .4772015    .6791433
        L2D. |   .2856943   .0385762     7.41   0.000     .2100231    .3613655
        L3D. |   .1419264   .0259157     5.48   0.000     .0910901    .1927627
      _trend |  -.0000581   .0001272    -0.46   0.648    -.0003077    .0001914
       _cons |   .0463977   .1073341     0.43   0.666     -.164149    .2569444
------------------------------------------------------------------------------
```

ARIMA MODEL

Conditions for determining p (number of lags) and q (moving average order) using acf and pacf

For ARIMA (p,d,0):

- ❖ The ACF is decaying and/or it follows a sine wave pattern.
- ❖ There are abnormal spikes in PACF at a certain lag (lag p), but none after that lag (lag p).

For ARIMA (0,d,q):

- ❖ The PACF is decaying and/or it follows a sine wave pattern.
- ❖ There are abnormal spikes in ACF at a certain lag (lag q), but none after that lag (lag q)

Determining p (number of lags) and q (moving average order) using acf and pacf in our model:

- ✓ P is 3 (we can observe 3 spikes in partial autocorrelation function graph)
- ✓ Q is 1 (we can observe 1 spikes in autocorrelation function graph)

```
. corrgram temp1
```

LAG	AC	PAC	Q	Prob>Q
1	-0.0623	-0.0623	5.6722	0.0172
2	-0.2421	-0.2473	91.512	0.0000
3	-0.0846	-0.1282	102	0.0000
4	-0.0505	-0.1417	105.74	0.0000
5	0.0287	-0.0527	106.94	0.0000
6	0.0416	-0.0268	109.48	0.0000
7	0.0122	-0.0135	109.7	0.0000
8	-0.0358	-0.0424	111.59	0.0000
9	0.0307	0.0316	112.97	0.0000
10	-0.0517	-0.0643	116.9	0.0000
11	0.0214	0.0225	117.58	0.0000
12	0.0139	-0.0105	117.86	0.0000
13	-0.0071	-0.0013	117.94	0.0000
14	-0.0166	-0.0242	118.35	0.0000
15	0.0387	0.0398	120.56	0.0000
16	0.0059	0.0038	120.61	0.0000
17	0.0216	0.0489	121.3	0.0000
18	-0.0080	0.0066	121.39	0.0000
19	-0.0159	0.0209	121.76	0.0000
20	-0.0046	-0.0005	121.79	0.0000
21	0.0036	0.0119	121.81	0.0000
22	0.0062	0.0030	121.87	0.0000
23	0.0397	0.0539	124.22	0.0000
24	0.0159	0.0287	124.59	0.0000
25	-0.0379	0.0014	126.72	0.0000
26	-0.0083	0.0050	126.83	0.0000
27	0.0183	0.0269	127.33	0.0000
28	0.0421	0.0492	129.97	0.0000
29	0.0236	0.0498	130.8	0.0000
30	-0.0185	0.0178	131.31	0.0000
31	0.0020	0.0429	131.31	0.0000
32	0.0238	0.0456	132.16	0.0000
33	0.0535	0.0939	136.44	0.0000
34	-0.0253	0.0183	137.39	0.0000
35	0.0159	0.0744	137.77	0.0000
36	-0.0232	0.0074	138.58	0.0000
37	-0.0210	0.0162	139.24	0.0000
38	0.0307	0.0323	140.66	0.0000
39	0.0309	0.0468	142.09	0.0000
40	0.0085	0.0325	142.2	0.0000

Dickey Fuller Test for Stationarity

Observations: With the trend term, we can see a slight difference in our results but it doesn't signify a trend. P- value is 0.0000 , therefore we can reject the null hypothesis and stationarity has been resolved.

```
. dfuller temp1, lags(3) trend regress

Augmented Dickey-Fuller test for unit root          Number of obs   =      1456

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -26.998            -3.960            -3.410            -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D.temp1      |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       temp1 |
         L1. |  -1.708098    .063267   -27.00   0.000    -1.832202   -1.583993
         LD. |   .5781724   .0514737    11.23   0.000     .4772015    .6791433
        L2D. |   .2856943   .0385762     7.41   0.000     .2100231    .3613655
        L3D. |   .1419264   .0259157     5.48   0.000     .0910901    .1927627
      _trend |  -.0000581   .0001272    -0.46   0.648    -.0003077    .0001914
       _cons |   .0463977   .1073341     0.43   0.666     -.164149    .2569444
------------------------------------------------------------------------------
```

ARIMA MODEL 1 (1,1,3)

```
ARIMA regression

Sample:  3/3/2013 - 2/28/2017                   Number of obs     =       1459
                                                Wald chi2(4)      =    1083.02
Log likelihood = -3131.268                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
     D.temp1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
temp1        |
       _cons |  -.0000419   .0000897    -0.47   0.640    -.0002176    .0001338
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.1091495   .0248034    -4.40   0.000    -.1577632   -.0605358
         L2. |  -.2578427   .0255704   -10.08   0.000    -.3079597   -.2077257
         L3. |  -.1284922   .0273979    -4.69   0.000    -.1821911   -.0747934
             |
          ma |
         L1. |         -1    .034016   -29.40   0.000     -1.06667   -.9333299
-------------+----------------------------------------------------------------
      /sigma |   2.063495          .        .       .            .           .
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)      df         AIC         BIC
-------------+---------------------------------------------------------------
           . |  1,459           .   -3131.268       5    6272.536    6298.964
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

For ARIMA(1,1,3)

- AIC: 6272.536
- BIC: 6298.964

From this we can observe that AR at L1, L2 and L3 are statistically significant in the model and when we put q,i.e, moving average as we see that MA is statistically significant.

ARIMA MODEL 2 (1,1,2)

ARIMA regression

Sample: 3/3/2013 - 2/28/2017 Number of obs = 1459
Wald chi2(3) = 99.82
Log likelihood = -3143.315 Prob > chi2 = 0.0000

D.temp1	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
temp1						
_cons	-.000043	.0001027	-0.42	0.676	-.0002443	.0001583
ARMA						
ar						
L1.	-.0773074	.0238815	-3.24	0.001	-.1241142	-.0305006
L2.	-.2472733	.0257632	-9.60	0.000	-.2977682	-.1967783
ma						
L1.	-1	40.68776	-0.02	0.980	-80.74655	78.74655
/sigma	2.080792	42.32981	0.05	0.480	0	85.0457

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

estat ic

kaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,459	.	-3143.315	5	6296.63	6323.057

Note: N=Obs used in calculating BIC; see [R] BIC note.

For ARIMA(1,1,2)

- AIC: 6296.63
- BIC: 6323.057

From this we can observe that AR at L1 and L2 are statistically significant in the model and when we put q,i.e, moving average as we see that MA is statistically non-significant

ARIMA MODEL 3 (1,1,1)

```
ARIMA regression

Sample:  3/3/2013 - 2/28/2017                   Number of obs     =       1459
                                                Wald chi2(2)      =       6.45
Log likelihood = -3189.186                      Prob > chi2       =     0.0397

------------------------------------------------------------------------------
             |                 OPG
     D.temp1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
temp1        |
       _cons |   -.000043   .0001309    -0.33   0.743    -.0002994    .0002135
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.0617421     .02433    -2.54   0.011     -.109428   -.0140562
             |
          ma |
         L1. |         -1   551.8482    -0.00   0.999    -1082.603    1080.603
-------------+----------------------------------------------------------------
      /sigma |   2.147692   592.5972     0.00   0.499            0    1163.617
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |      1,459         .  -3189.186       4    6386.372   6407.514
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

For ARIMA(1,1,1)

- AIC: 6386.372
- BIC: 6407.514

From this we can observe that AR at L1 is statistically significant even at 5% and 10% confidence but non-significant at 1% confidence interval in the model and when we put q,i.e, moving average as we see that MA is statistically non-significant

Summary for all three ARIMA models-

From this we can observe that AR at L1 is statistically significant in all the three models whereas when we put q, i.e, moving average as 1, 2 respectively, we see that ma is not statistically significant and if q is 3, ma is statistically significant.

We can see that the model 1 (ARIMA 1,1,3) has the lowest bic and aic and hence it is the best model.

```
. wntestq temp1, lag(8)

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =   111.5856
 Prob > chi2(8)            =     0.0000
```

```
. wntestq temp1, lag(24)

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =   124.5925
 Prob > chi2(24)           =     0.0000
```

```
. wntestq temp1, lag(16)

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =   120.6087
 Prob > chi2(16)           =     0.0000
```

Determining lags

```
. varsoc TEMP PM10 SO2 NO2 CO O3 PM25

   Selection-order criteria
   Sample:  3/5/2013 - 2/28/2017                  Number of obs     =      1457
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-51428.5				1.1e+22	70.6047	70.6142	70.6301
1	-46756.7	9343.7	49	0.000	1.9e+19	64.259	64.3348	64.4621*
2	-46611.9	289.61	49	0.000	1.7e+19	64.1275	64.2696	64.5083
3	-46503.3	217.1	49	0.000	1.5e+19	64.0458	64.2541*	64.6043
4	-46437.9	130.82*	49	0.000	1.5e+19*	64.0232*	64.2979	64.7595

```
   Endogenous:  TEMP PM10 SO2 NO2 CO O3 PM25
    Exogenous:  _cons
```

The LR ,AIC and FPE have chosen a model with four lags, whereas SBIC and HQIC have selected a model with three and two lags respectively.

We will consider 4 lags for the following tests.

Johansen test for cointegration

```
. vecrank TEMP PM10 SO2 NO2 CO O3 PM25, lag(4)

                    Johansen tests for cointegration
Trend: constant                                   Number of obs =    1457
Sample:  3/5/2013 - 2/28/2017                              Lags =       4
```

maximum rank	parms	LL	eigenvalue	trace statistic	5% critical value
0	154	-1997.8357	.	149.3540	124.24
1	167	-1975.9786	0.48435	105.6399	94.15
2	178	-1956.6199	0.44380	66.9224*	68.52
3	187	-1939.2922	0.40849	32.2671	47.21
4	194	-1932.5083	0.18582	18.6992	29.68
5	199	-1926.6669	0.16223	7.0165	15.41
6	202	-1923.2226	0.09911	0.1278	3.76
7	203	-1923.1587	0.00193		

We cannot reject the null hypothesis at rank 2 and conclude that there are 2 cointegration equations.

Fitting VECMs with Johansen's normalization

```
. vec TEMP PM10 SO2 NO2 CO O3 PM25, lag(4) rank(2)

Vector error-correction model

Sample:  3/5/2013 - 2/28/2017                   Number of obs     =      1,457
                                                AIC               =   64.33474
Log likelihood = -46689.86                      HQIC              =   64.57558
Det(Sigma_ml)  =  1.61e+19                      SBIC              =   64.98029

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_TEMP               24    1.95638   0.1819   318.4951   0.0000
D_PM10               24    65.3272   0.2767    547.929   0.0000
D_SO2                24    13.8399   0.2840   568.0612   0.0000
D_NO2                24    20.4393   0.2630   511.0171   0.0000
D_CO                 24    762.524   0.2312   430.6819   0.0000
D_O3                 24     23.685   0.2205   405.0371   0.0000
D_PM25               24    59.5877   0.2719    534.734   0.0000
----------------------------------------------------------------
```

Having determined that there exists cointegrating equation, we now want to estimate the parameter of cointegrating VECM for these series by using vec.

```
-------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
D_TEMP       |
         _ce1|
          L1.|  -.0123632    .0073438    -1.68   0.092    -.0267568    .0020304
             |
         _ce2|
          L1.|  -.0011737    .0007787    -1.51   0.132    -.0026999    .0003525
             |
         TEMP|
          LD.|  -.0857526    .0279187    -3.07   0.002    -.1404722    -.031033
         L2D.|  -.2622206    .0275326    -9.52   0.000    -.3161835   -.2082578
         L3D.|  -.1017516    .0272369    -3.74   0.000    -.1551349   -.0483684
             |
         PM10|
          LD.|  -.0020178    .0020466    -0.99   0.324    -.0060291    .0019935
         L2D.|  -.0001975    .0020076    -0.10   0.922    -.0041323    .0037373
         L3D.|  -.0069459    .0019613    -3.54   0.000      -.01079   -.0031018
             |
          SO2|
          LD.|   .0106315     .004948     2.15   0.032     .0009337    .0203293
         L2D.|    .001537    .0049007     0.31   0.754    -.0080681    .0111422
         L3D.|   .0010929    .0046989     0.23   0.816    -.0081167    .0103025
             |
          NO2|
          LD.|   .0182775    .0039627     4.61   0.000     .0105107    .0260443
         L2D.|   .0025712    .0040525     0.63   0.526    -.0053716     .010514
         L3D.|   .0136983    .0040028     3.42   0.001      .005853    .0215436
             |
           CO|
```

Fitting VECMs with Johansen's normalization

Cointegrating equations

Equation	Parms	chi2	P>chi2
_ce1	5	317.8175	0.0000
_ce2	5	224.9736	0.0000

We fit a VECM with 2 cointegrating equations and 5 lags.
From the table on the right with the restrictions imposed we can

beta-hat=1 v-hat= -101.8777

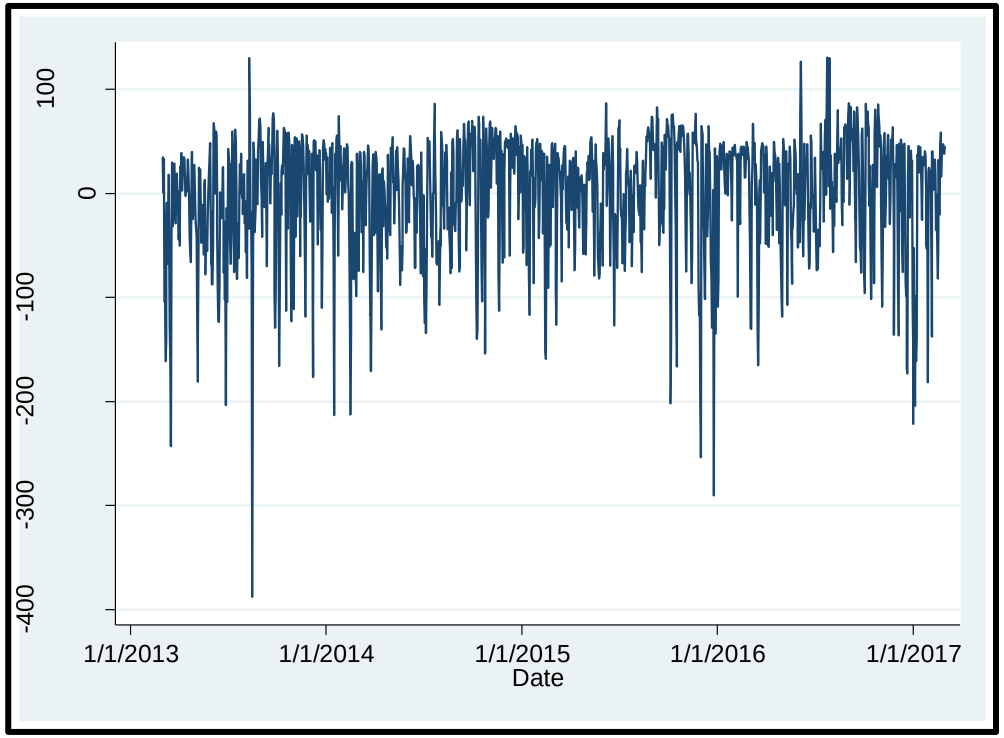
The output indicates that the model fits well and the coefficients in the cointegrating equation are statistically significant, as are the adjustment parameters.

Identification: beta is exactly identified

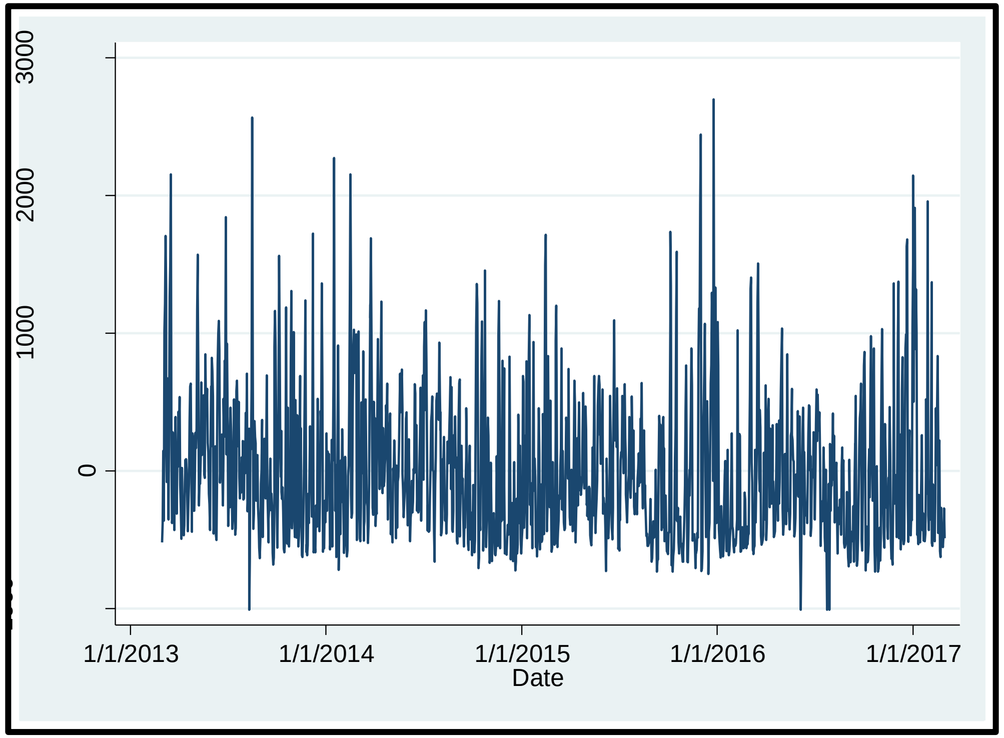
Johansen normalization restrictions imposed

beta	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_ce1						
TEMP	1	.	.	.	.	.
PM10	-2.78e-17	.	.	.	.	.
SO2	-.0765576	.1765959	-0.43	0.665	-.4226792	.269564
NO2	-.3843875	.148734	-2.58	0.010	-.6759008	-.0928741
CO	.0106437	.0048309	2.20	0.028	.0011754	.0201121
O3	-.8150983	.0798111	-10.21	0.000	-.9715252	-.6586714
PM25	-.7285311	.0770657	-9.45	0.000	-.8795772	-.577485
_cons	101.8777	.	.	.	.	.
_ce2						
TEMP	-8.88e-16	.	.	.	.	.
PM10	1	.	.	.	.	.
SO2	3.985261	1.645981	2.42	0.015	.7591983	7.211324
NO2	3.169608	1.386291	2.29	0.022	.4525272	5.886689
CO	-.0741367	.0450267	-1.65	0.100	-.1623874	.014114
O3	5.041204	.7438881	6.78	0.000	3.58321	6.499198
PM25	5.588243	.7182994	7.78	0.000	4.180402	6.996084
_cons	-1009.207	.	.	.	.	.

Predicting the cointegrating equations and graphing them



Cointegration Equation 1



Cointegration Equation 2

Testing for serial correlation in the residuals

At all lags, reject the null hypothesis as the value is not more than 0.01, 0.05 or 0.1. Therefore there is autocorrelation.

```
. veclmar, mlag(4)

   Lagrange-multiplier test
```

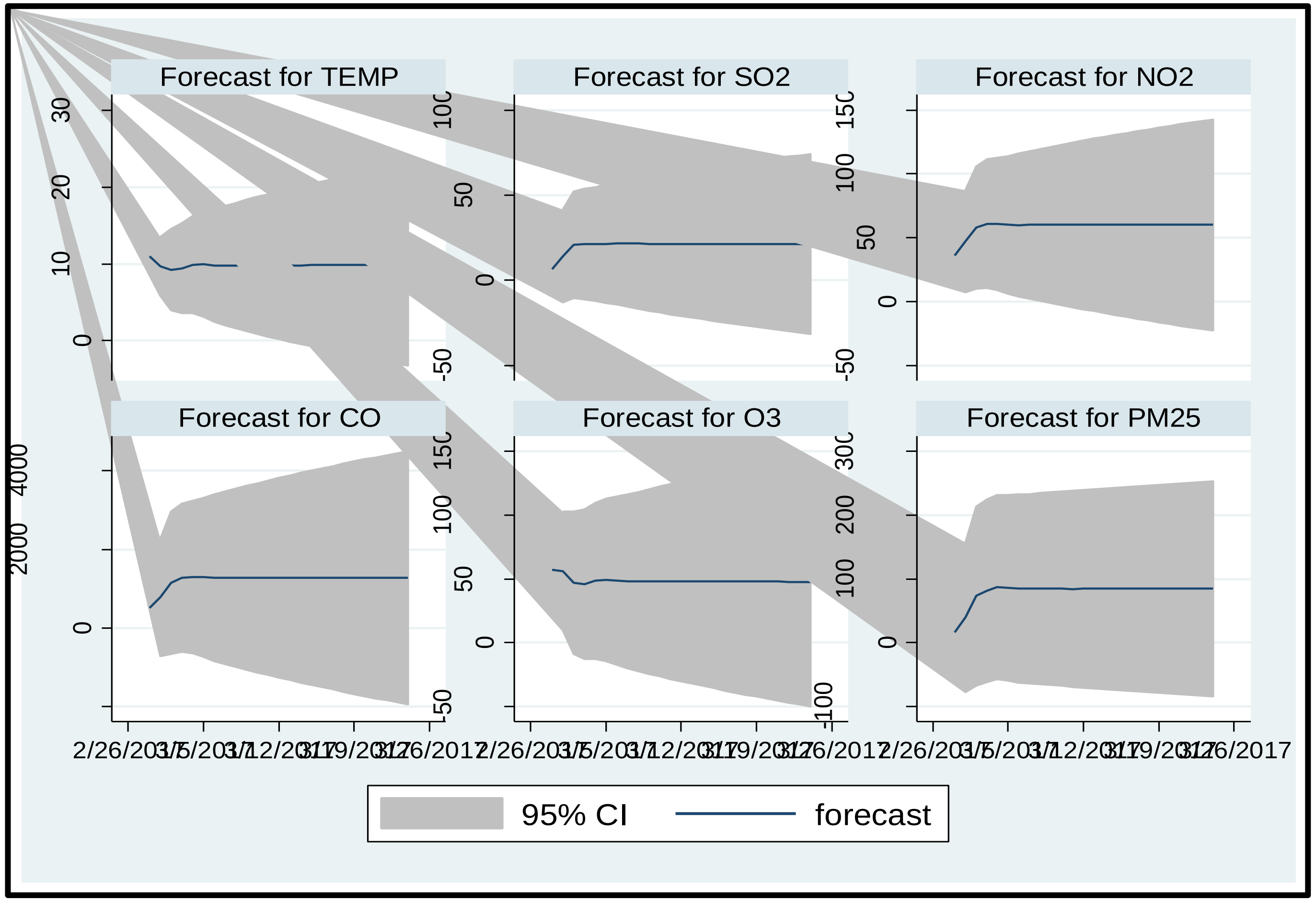
lag	chi2	df	Prob > chi2
1	259.2926	49	0.00000
2	300.4627	49	0.00000
3	374.2985	49	0.00000
4	383.5111	49	0.00000

H0: no autocorrelation at lag order

Forecasting with VECMs

We use fcast compute to obtain dynamic forecasts of the levels and fcast graph to graph these dynamic forecasts, along with their asymptotic confidence intervals.

As for temperature we cannot see any gradual change, rather it seems consistant.



Thank you