Report

DNSC6219 – Time Series Analysis

Alexis Yang, Chengshu Yang, Qunzhe Ding, Seungheon Han

1. Introduction and Overview.

The dataset we chose to use is weather forecasting for Indian climate. This dataset provides daily data from 1st January 2013 to 24th April 2017 in the city of Delhi, India. The 4 parameters here are meantemp, humidity, wind_speed, meanpressure (Pressure reading of weather). Our interest is to select the best forecasting model among the below time-series models. The link of the dataset is (www.kaggle.com/sumanthvrao/daily-climate-time-series-data).

2. Univariate Time-series models.

Fitting the univariate model with the variable MEANTEMP

series:

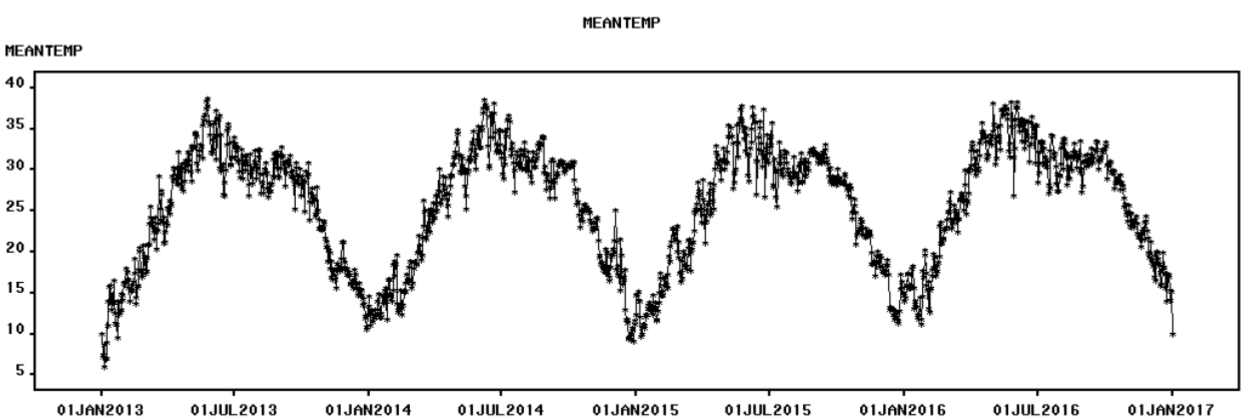

Figure1

As shown in the figure1, the series is characterized by the yearly-based seasonal pattern created by the daily data. The variance of the seasonal patterns look similar.

2.1 Deterministic Time Series Models (Seasonal Dummies and Trend, Cyclical Trend) and Error model.

Using forecast horizon: 12, Hold-out Sample: 20

Seasonal Dummies and Trend:

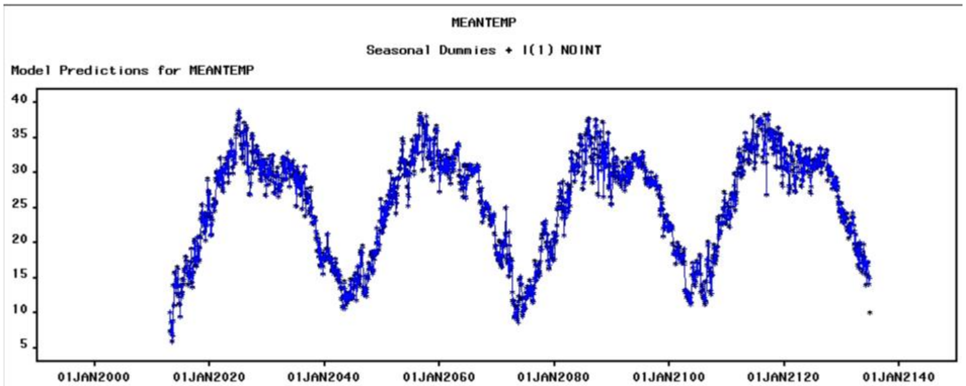

Figure2-1

Model Viewer

Parameter Estimates
MEANTEMP
Seasonal Dummies + I(1)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	-0.03077	0.1522	-0.2022	0.8448
Seasonal Dummy 1	0.07801	0.2152	0.3625	0.7264
Seasonal Dummy 2	0.19004	0.2148	0.8849	0.4020
Seasonal Dummy 3	0.07725	0.2152	0.3590	0.7289
Seasonal Dummy 4	0.18663	0.2152	0.8672	0.4111
Seasonal Dummy 5	-0.03049	0.2152	-0.1417	0.8908
Seasonal Dummy 6	-0.14705	0.2152	-0.6833	0.5137
Seasonal Dummy 7	0.16260	0.2152	0.7556	0.4716
Seasonal Dummy 8	-0.28595	0.2152	-1.3287	0.2206
Seasonal Dummy 9	0.15285	0.2152	0.7103	0.4977
Seasonal Dummy 10	-0.01412	0.2152	-0.0656	0.9493
Seasonal Dummy 11	0.08072	0.2152	0.3751	0.7173
Model Variance (sigma squared)	2.77873	.	.	.

Figure2-2

Statistics of Fit
MEANTEMP
Seasonal Dummies + I(1)

Statistic of Fit	Value
Mean Square Error	3.86477
Root Mean Square Error	1.96590
Mean Absolute Percent Error	10.74431
Mean Absolute Error	1.60737
R-Square	0.155

Figure2-3

According to the figures (2-1, 2, 3) above, we applied seasonal dummies regressor and first difference trend model, the model prediction for MeanTemp fits well and the mean absolute percent error is 10.74431, the p value of monthly dummy variables are greater than 0.05 and R-Square is 0.155. Therefore, this is not a desired model among others listed below.

Cyclical model of Meantemp

Identify the 6 harmonics (for purposes of parsimony) with the highest amplitudes to include in a cyclical trend model.

i	Periods
4	365.50
8	182.75
12	121.83
3	487.33
1	1462.00
7	208.86

Table1

Plot the periodogram:

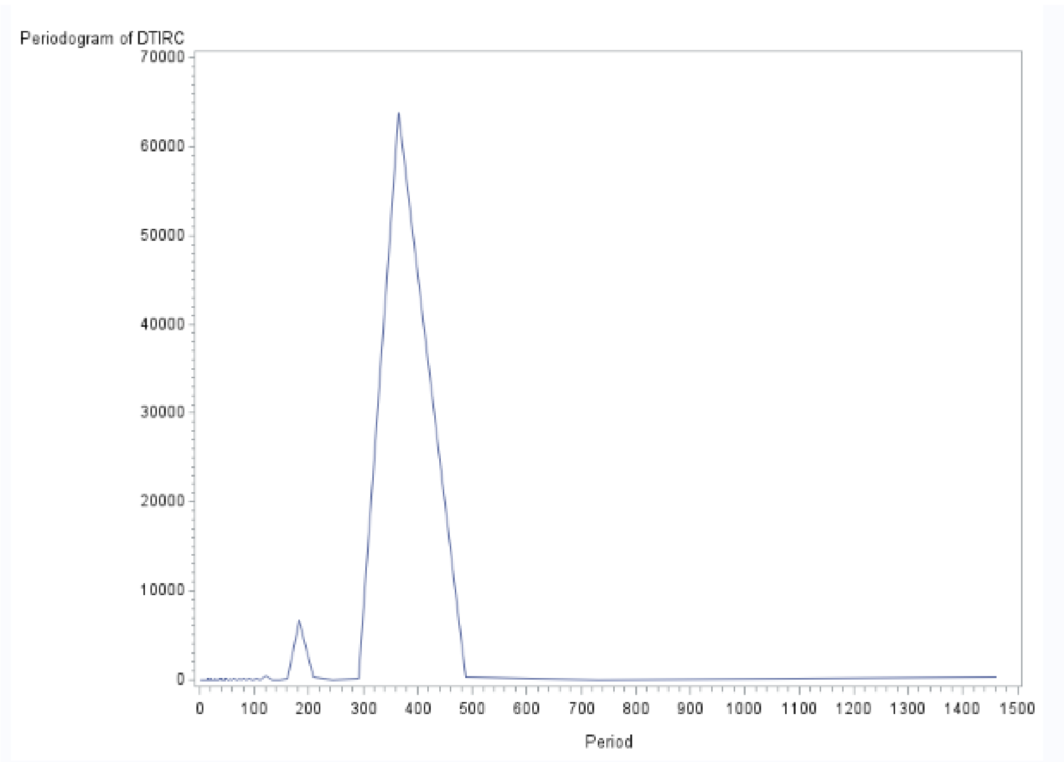

Figure3-1

Create the necessary sine and cosine pairs we have identified from the periodogram. Plot of actual versus predicted values.

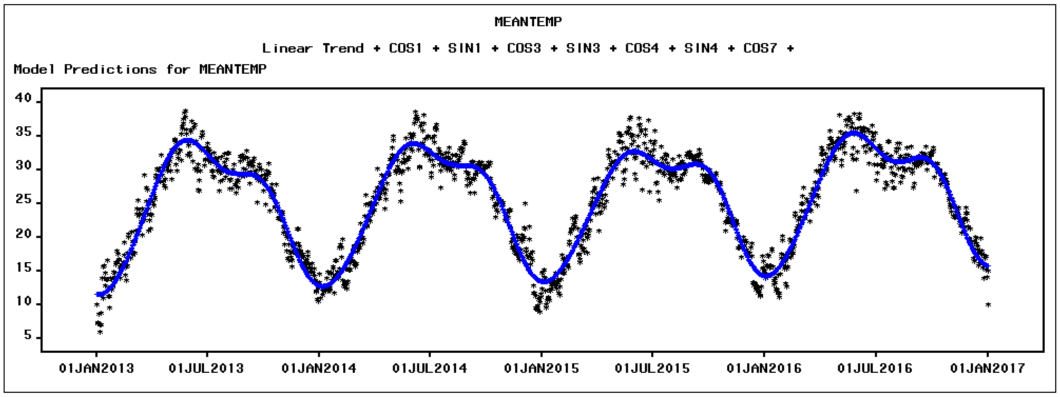

Figure3-2

Parameter Estimates
MEANTEMP
Linear Trend + COS1 + SIN1 + COS3 + SIN3 + COS4 + SIN4 + COS7 + SIN7 + COS8 + SIN8 + COS12 + SIN12

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Linear Trend	0.00289	0.000267	10.8524	<.0001
COS1	0.50641	0.0771	6.5693	0.0006
SIN1	0.52037	0.1452	3.5841	0.0116
COS3	-0.18005	0.0771	-2.3364	0.0581
SIN3	0.66265	0.0860	7.7071	0.0003
COS4	-9.33400	0.0770	-121.1572	<.0001
SIN4	-0.35600	0.0815	-4.3665	0.0047
COS7	-0.21393	0.0769	-2.7803	0.0320
SIN7	-0.46482	0.0775	-5.9941	0.0010
COS8	-2.58925	0.0769	-33.6691	<.0001
SIN8	-1.50509	0.0771	-19.5188	<.0001
COS12	-0.01975	0.0767	-0.2575	0.8054
SIN12	0.74354	0.0764	9.7284	<.0001
Model Variance (sigma squared)	4.18692	.	.	.

Figure3-3

Statistic of Fit	Value
Mean Square Error	3.62865
Root Mean Square Error	1.90490
Mean Absolute Percent Error	10.08953
Mean Absolute Error	1.51104
R-Square	0.206

Figure3-4

Model variance:

Seasonal model (Figure2-2): 2.77873

Cyclical model (Figure3-3): 4.18692

RMSE:

Seasonal model (Figure2-3): 1.9659

Cyclical model (Figure3-4): 1.9049

From the above plots from cyclical model and seasonal dummies and trend model, since there are too many observations, we are not reasonably sure to infer which model fits the data better. However, the seasonal model has a lower model variance and the cyclical model has a lower root mean square error. P-values for the cyclical model are almost all less than 0.05 except cos12. Compared with the seasonal model, we are reasonably sure that the cyclical model fits the data better.

Error model of Meantemp

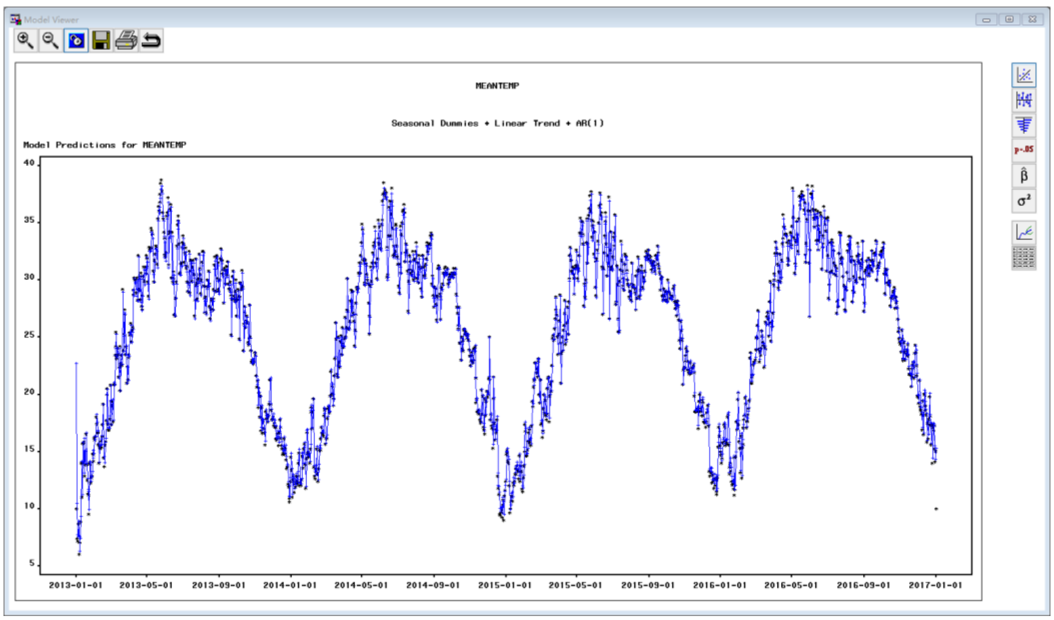

Figure4-1

MEANTEMP

Linear Trend + Seasonal Dummies + AR(1)

Statistic of Fit	Value
Mean Square Error	4.60044
Root Mean Square Error	2.14486
Mean Absolute Percent Error	11.57050
Mean Absolute Error	1.79954
R-Square	0.077

Figure4-2

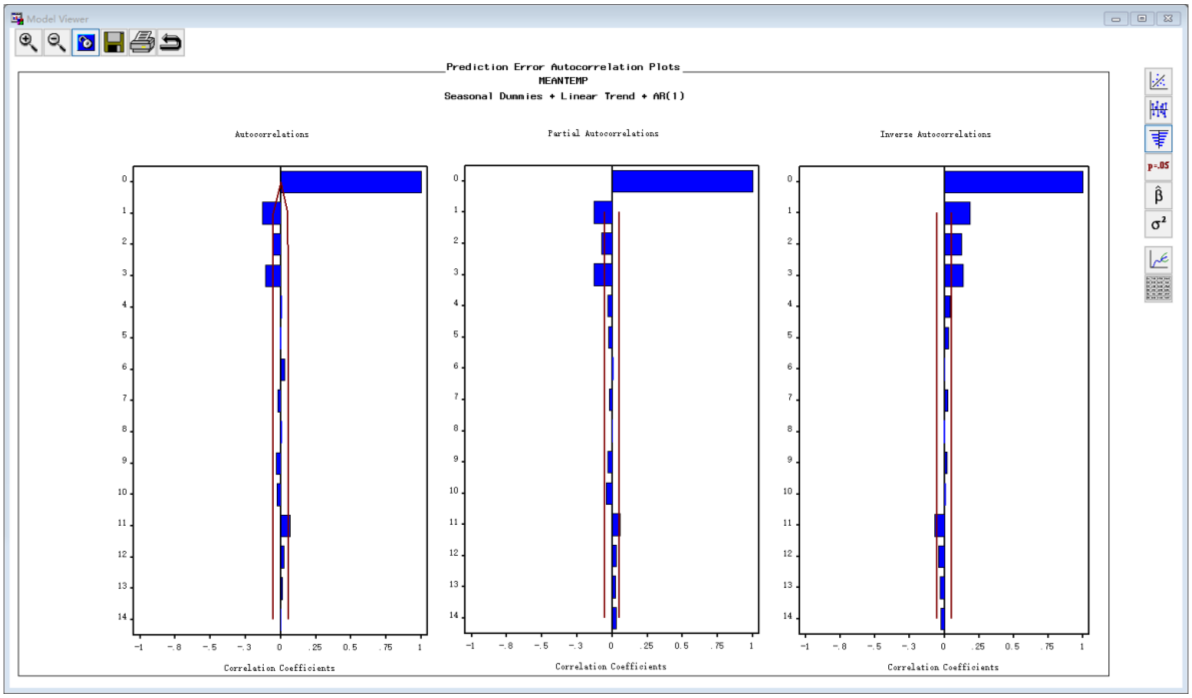

Figure4-3

ACF of the error model cuts off to zero after lag 3 and PACF decays exponentially sending the PAC to 0 after lag 3.

Parameter Estimates
MEANTEMP
Seasonal Dummies + Linear Trend + AR(1)

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Intercept	22.60908	3.2406	6.9769	<.0001
Autoregressive, Lag 1	0.97474	0.0059	164.4969	<.0001
Seasonal Dummy 1	-0.10095	0.1082	-0.9329	0.3657
Seasonal Dummy 2	0.03810	0.1397	0.2728	0.7887
Seasonal Dummy 3	0.08453	0.1529	0.5527	0.5886
Seasonal Dummy 4	0.20496	0.1529	1.3404	0.2001
Seasonal Dummy 5	0.13702	0.1397	0.9811	0.3421
Seasonal Dummy 6	0.10782	0.1082	0.9963	0.3349
Linear Trend	0.00325	0.0039	0.8435	0.4122
Model Variance (sigma squared)	2.73739	.	.	.

Figure4-4

According to the output, the root mean square error of 2.145 is pretty small, the mean absolute percent error is acceptably small (Figure4-2). The fit model of error method fits pretty well. However, the p-values of seasonal dummy are all above 0.05 (Figure4-4),which represents that the model is not very desired as we wish.

2.2 ARIMA models (with seasonal ARIMA components if relevant)

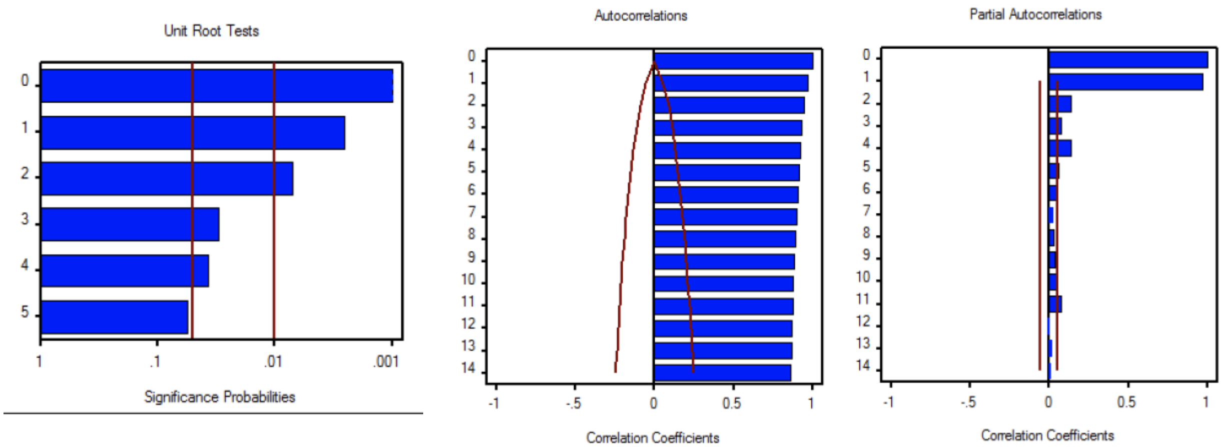

Figure5-1

Regarding the stationarity of the original series (Figure5-1), the Unit Root Test shows that the series turns into non-stationary from lag 5. The ACF of the series which is decaying very slowly also implies the series is non-stationary.

So, we can take first-difference:

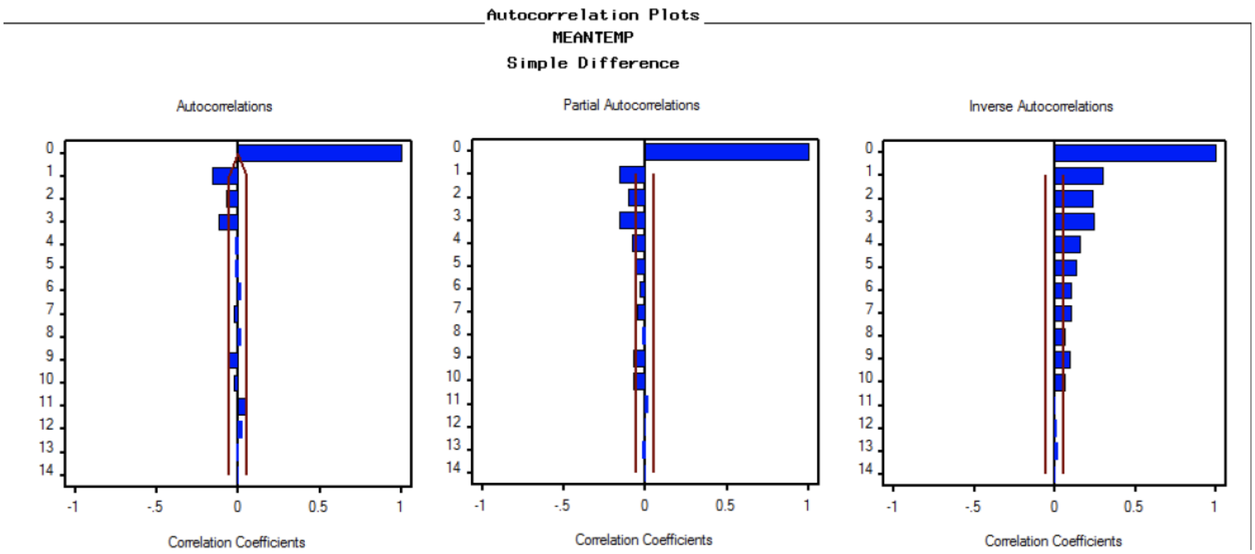

Figure5-2

ACF of the simple difference model cuts off to zero after lag 3 and PACF decays exponentially sending the PAC to 0 after lag 4 (Figure5-2). Based on these features, we can decide to fit the Integrated Moving Average model with the order of 3.

IMA(1,3) with 20 hold-out sample:

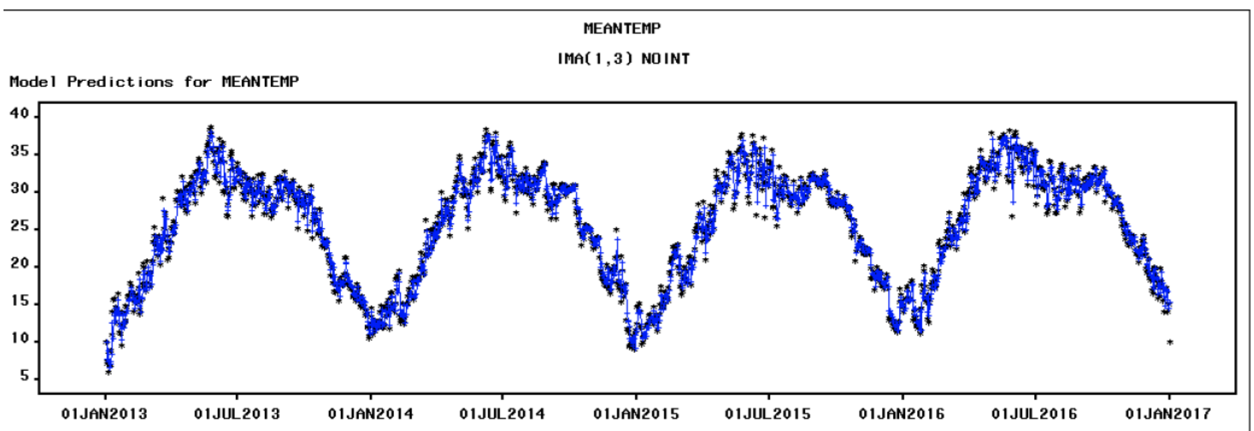

Figure5-3

The estimated IMA model seems to well predict the original series (Figure5-3).

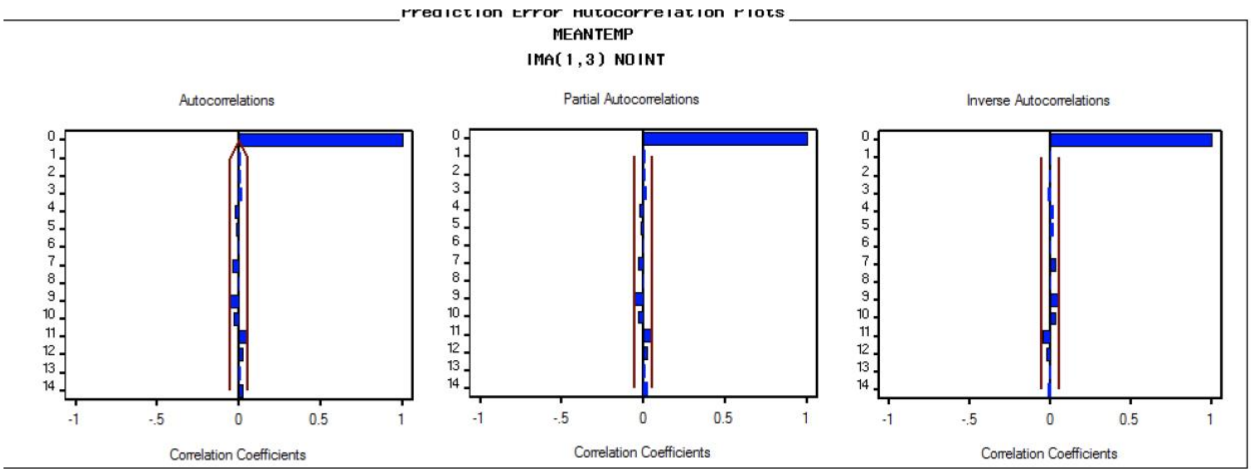

Figure5-4

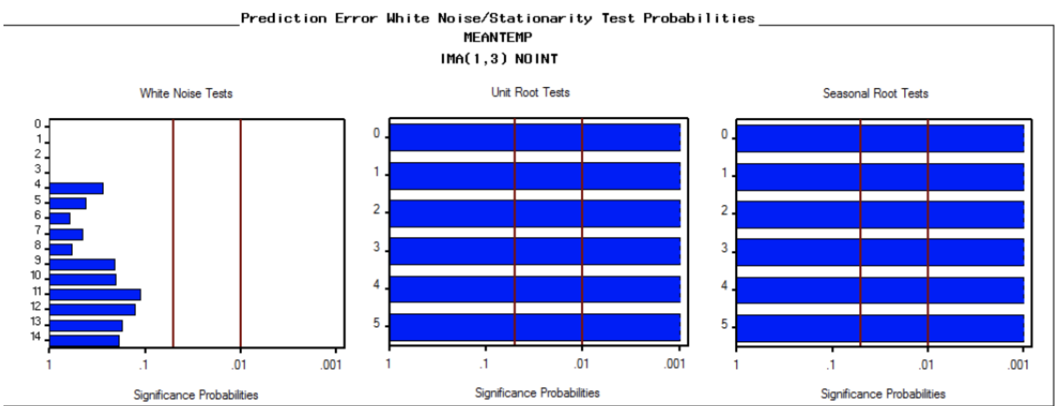

Figure5-5

According to the ACF, PACF and IACF of the residuals (Figure5-4) staying outside the bounds, we can be reasonably sure that the residuals of the estimated model are not correlated with each other. The Ljung-Box test makes it clearer that the residuals are white noise terms.

Beside, the Unit Root Test (Figure5-5) demonstrates the stationarity of the residuals.

Parameter Estimates
MEANTEMP
IMA(1,3) NOINT

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Moving Average, Lag 1	0.22134	0.0261	8.4909	<.0001
Moving Average, Lag 2	0.12953	0.0265	4.8861	0.0001
Moving Average, Lag 3	0.15262	0.0261	5.8415	<.0001
Model Variance (sigma squared)	2.57226	.	.	.

Figure5-6

The three MA coefficients are all statistically significant with the p-values smaller than 0.05 and the goodness-of-fit of the model can be considered fine based on its low model variance which is around 2.57 (Figure5-6).

Statistics of Fit
MEANTEMP
IMA(1,3) NOINT

Statistic of Fit	Value
Mean Square Error	3.71560
Root Mean Square Error	1.92759
Mean Absolute Percent Error	10.11047
Mean Absolute Error	1.47646
R-Square	0.187

Figure5-7

Lastly, the measures in the **Figure5-7** demonstrate the reasonable goodness-of-fit of the IMA(1,3) model on the basis of the low values.

2.3 Comparison of models (in terms of fit and validation)

MODEL	Model Variance	MAPE	MAE	P-values
Seasonal Dummies	**2.7934**	**10.74431**	**1.6074**	**Seasonal dummies > 0.05, not statistically significant**
Cyclical	**4.1869**	**10.08953**	**1.5110**	**Every coefficient has the p-value < 0.05 excepting cos12**
Error	**2.7374**	**11.5705**	**1.7995**	**Seasonal dummies > 0.05, not statistically significant**
IMA(1,3)	**2.5723**	**10.1105**	**1.4765**	**Every MA coefficient**

				has the p-values < 0.05

Table2

- **According to the table2, the IMA(1,3) model has the lowest values in terms of Model Variance and Mean Absolute Error. Despite the fact that the cyclical model has slightly lower MAPE compared to the IMA model, the difference is so small that we can select the IMA(1,3) model as the best one when it comes to the model's goodness-of-fit.**

- **Validation can be measured based on the models' p-values. Cyclical and MA are the only models whose t-test p-values are statistically significant.**