12/5/2018

Time series and Forecasting

Case Study

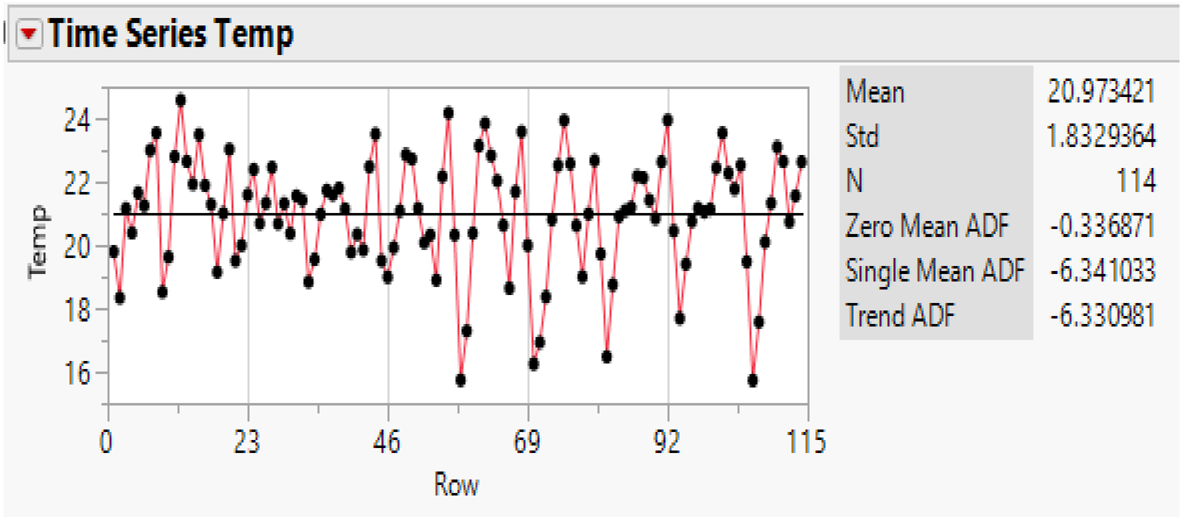
Jay Bhanushali

The software used in this report is JMP

1)

Holt-Winter's Forecasting method:

The Time-series output from JMP is as follows:



From the Time-series plot, we can see that there is no significant amplitude increase as it is almost constant. Thus, we use the Holt-Winter's additive method for our problem.

The ACF and PACF plots are as follows:

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.4629	25.0810	<.0001*	1	0.4629
2	-0.1834	29.0521	<.0001*	2	-0.5062
3	-0.2959	39.4862	<.0001*	3	0.1134
4	-0.1693	42.9312	<.0001*	4	-0.1834
5	-0.0825	43.7566	<.0001*	5	-0.0450
6	-0.0543	44.1171	<.0001*	6	-0.0852
7	0.0068	44.1227	<.0001*	7	0.0245
8	-0.0745	44.8151	<.0001*	8	-0.2762
9	-0.2370	51.8879	<.0001*	9	-0.1602
10	-0.2481	59.7173	<.0001*	10	-0.1977
11	0.1972	64.7096	<.0001*	11	0.4800
12	0.6426	118.252	<.0001*	12	0.2314
13	0.3356	133.000	<.0001*	13	-0.2920
14	-0.1143	134.729	<.0001*	14	0.2022
15	-0.2192	141.147	<.0001*	15	-0.0958
16	-0.1200	143.090	<.0001*	16	0.1065
17	-0.0405	143.313	<.0001*	17	0.0375
18	-0.0204	143.371	<.0001*	18	-0.0863
19	-0.0115	143.389	<.0001*	19	-0.1206
20	-0.1070	145.001	<.0001*	20	-0.0354
21	-0.2299	152.516	<.0001*	21	-0.0439
22	-0.2243	159.749	<.0001*	22	0.0789
23	0.1302	162.211	<.0001*	23	-0.0061
24	0.4748	195.336	<.0001*	24	-0.0456
25	0.2702	206.185	<.0001*	25	-0.0103

The model summary and the parameter's estimate are as follows:

Model: Winters Method (Additive)

Model Summary

DF	98	Stable	Yes
Sum of Squared Errors	167.498463	Invertible	No
Variance Estimate	1.70916799		
Standard Deviation	1.30735152		
Akaike's 'A' Information Criterion	349.518438		
Schwarz's Bayesian Criterion	357.3638		
RSquare	0.45996215		
RSquare Adj	0.44894097		
MAPE	5.14585419		
MAE	1.06607863		
-2LogLikelihood	343.518438		

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	1.74478e-8	1.8354e-7	0.10	0.9245
Trend Smoothing Weight	0.00009885	0.0012006	0.08	0.9346
Seasonal Smoothing Weight	0.76083326	0.1390828	5.47	<.0001*

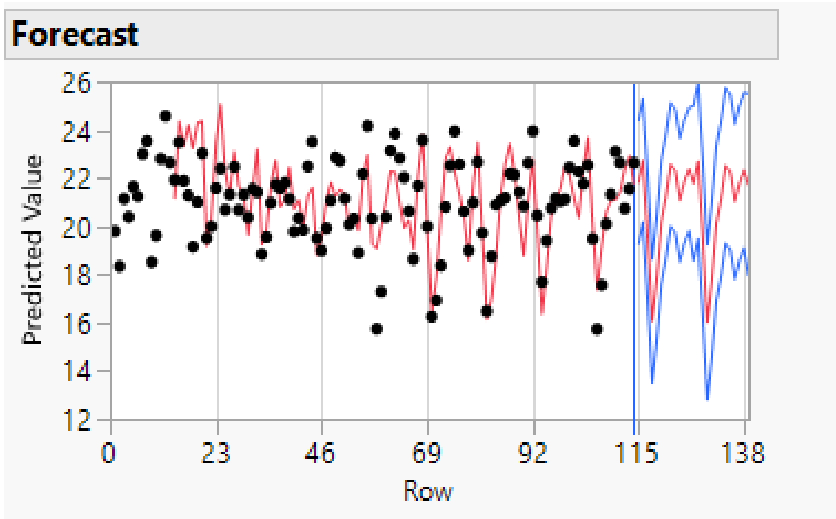
The level smoothing weight and Trend smoothing weight are not significant which can be seen from the parameter estimates. This validates the fact that there is no significant increase in the level and trend of the time-series.

The accuracy measures are as follows:

Model	DF	Variance	AIC	SBC	RSquare	Weights	.2 .4 .6 .8	MAPE	MAE
Winters Method (Additive)	98	1.709168	349.51844	357.36380	0.460	1.000000		5.145854	1.066079

The output from JMP gives just the optimal model. Thus, we can conclude that this model is the best for the Temperature time-series.

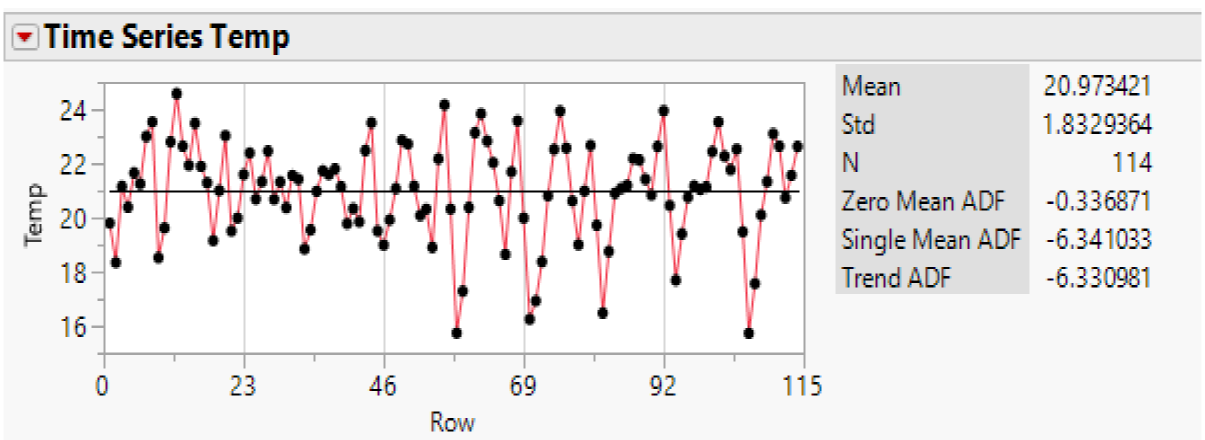
The forecast plot is as follows:



2)

Modeling ARIMA model on the temperature time-series:

The time-series plot is as follows:



As we can see from the time-series plot from above, this is stationary seasonal time-series.

The ACF and PACF plots are as follows:

Time Series Basic Diagnostics

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.4629		25.0810	<.0001*	1	0.4629	
2	-0.1834		29.0521	<.0001*	2	-0.5062	
3	-0.2959		39.4862	<.0001*	3	0.1134	
4	-0.1693		42.9312	<.0001*	4	-0.1834	
5	-0.0825		43.7566	<.0001*	5	-0.0450	
6	-0.0543		44.1171	<.0001*	6	-0.0852	
7	0.0068		44.1227	<.0001*	7	0.0245	
8	-0.0745		44.8151	<.0001*	8	-0.2762	
9	-0.2370		51.8879	<.0001*	9	-0.1602	
10	-0.2481		59.7173	<.0001*	10	-0.1977	
11	0.1972		64.7096	<.0001*	11	0.4800	
12	0.6426		118.252	<.0001*	12	0.2314	
13	0.3356		133.000	<.0001*	13	-0.2920	
14	-0.1143		134.729	<.0001*	14	0.2022	
15	-0.2192		141.147	<.0001*	15	-0.0958	
16	-0.1200		143.090	<.0001*	16	0.1065	
17	-0.0405		143.313	<.0001*	17	0.0375	
18	-0.0204		143.371	<.0001*	18	-0.0863	
19	-0.0115		143.389	<.0001*	19	-0.1206	
20	-0.1070		145.001	<.0001*	20	-0.0354	
21	-0.2299		152.516	<.0001*	21	-0.0439	
22	-0.2243		159.749	<.0001*	22	0.0789	
23	0.1302		162.211	<.0001*	23	-0.0061	
24	0.4748		195.336	<.0001*	24	-0.0456	
25	0.2702		206.185	<.0001*	25	-0.0103	

Both ACF and PACF plots have some season component which are significant in it which has to be modeled out. From the plots above, we can see a cut-off pattern on both ACF and PACF plots, suggesting the best model might be an ARMA model with seasonal component.

Now we apply a seasonal difference of order 1 to see if the seasonal components in ACF and PACF are still significant.

The ACF and PACF plots are given as:

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.5696		34.0795	<.0001*	1	0.5696	
2	-0.0279		34.1620	<.0001*	2	-0.5216	
3	-0.2638		41.6212	<.0001*	3	0.1168	
4	-0.2588		48.8710	<.0001*	4	-0.2020	
5	-0.1389		50.9823	<.0001*	5	0.0714	
6	0.0655		51.4570	<.0001*	6	0.1005	
7	0.2107		56.4132	<.0001*	7	0.0417	
8	0.1462		58.8260	<.0001*	8	-0.0750	
9	-0.0405		59.0131	<.0001*	9	-0.0495	
10	-0.2025		63.7406	<.0001*	10	-0.1238	
11	-0.2261		69.7001	<.0001*	11	0.0034	
12	-0.1028		70.9453	<.0001*	12	0.0065	
13	0.0635		71.4266	<.0001*	13	0.0240	
14	0.0738		72.0838	<.0001*	14	-0.1730	
15	-0.0193		72.1294	<.0001*	15	0.0140	
16	-0.0747		72.8186	<.0001*	16	-0.0378	
17	-0.0742		73.5054	<.0001*	17	0.0226	
18	-0.0105		73.5194	<.0001*	18	0.0663	
19	0.0686		74.1205	<.0001*	19	-0.0177	
20	0.0616		74.6110	<.0001*	20	-0.0946	
21	-0.0302		74.7307	<.0001*	21	-0.0451	
22	-0.1029		76.1351	<.0001*	22	-0.0368	
23	-0.1654		79.8104	<.0001*	23	-0.1430	
24	-0.1931		84.8797	<.0001*	24	-0.0809	
25	-0.1284		87.1514	<.0001*	25	-0.0914	

Now we can see that the seasonal components are no more significant. The cut-off pattern can be seen in both the plots, suggesting we should try an ARMA model after the seasonal differencing:

We consider 2 candidate models for comparison:

1) ARIMA (2, 0, 1) x $(0, 1, 0)_{12}$
2) ARIMA (2, 0, 0) x $(0, 1, 0)_{12}$

ARIMA (2, 0, 1) x $(0, 1, 0)_{12}$

The model summary and the parameter estimates are given below:

Model: Seasonal ARIMA(2, 0, 1)(0, 1, 0)12

Model Summary

DF	98	Stable	Yes
Sum of Squared Errors	77.6823524	Invertible	Yes
Variance Estimate	0.79267707		
Standard Deviation	0.89032413		
Akaike's 'A' Information Criterion	270.968381		
Schwarz's Bayesian Criterion	281.468272		
RSquare	0.75307253		
RSquare Adj	0.74551353		
MAPE	3.43939289		
MAE	0.70569686		
-2LogLikelihood	262.968381		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	0.6254392	0.1626991	3.84	0.0002*
AR1,2	1	2	-0.3947507	0.1380724	-2.86	0.0052*
MA1,1	1	1	-0.4005426	0.1719332	-2.33	0.0219*
Intercept	1	0	-0.0451508	0.1554803	-0.29	0.7721

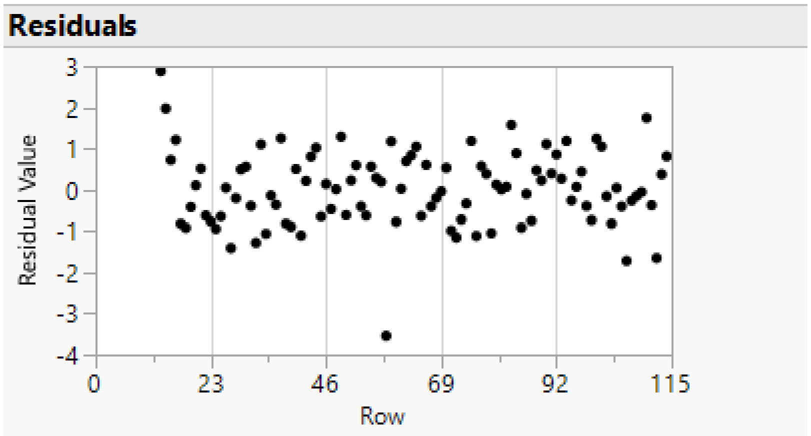
Constant Estimate	Mu
-0.034735	-0.0451508

As we can see, all the parameters are significant except the MA1 component which is not as significant as others, suggesting this model is still considered as valid.

The accuracy measures are given as:

Model	DF	Variance	AIC	SBC	RSquare	MAPE	MAE
Seasonal ARIMA(2, 0, 1)(0, 1, 0)12	98	0.7926771	270.96838	281.46827	0.753	3.439393	0.705697

Residual analysis:



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0214		0.0479	0.8268	1	0.0214	
2	-0.0203		0.0918	0.9551	2	-0.0208	
3	0.0149		0.1156	0.9899	3	0.0158	
4	-0.0495		0.3811	0.9840	4	-0.0507	
5	-0.0758		1.0093	0.9618	5	-0.0732	
6	-0.0374		1.1636	0.9786	6	-0.0369	
7	0.1489		3.6379	0.8204	7	0.1502	
8	-0.0167		3.6695	0.8856	8	-0.0251	
9	-0.0357		3.8149	0.9232	9	-0.0370	
10	-0.0997		4.9603	0.8938	10	-0.1173	
11	-0.0665		5.4763	0.9059	11	-0.0545	
12	-0.1442		7.9268	0.7908	12	-0.1309	
13	0.0774		8.6400	0.7996	13	0.0965	
14	0.0115		8.6559	0.8524	14	-0.0357	
15	-0.0482		8.9397	0.8807	15	-0.0588	
16	0.0066		8.9451	0.9157	16	-0.0232	
17	-0.0328		9.0790	0.9378	17	-0.0208	
18	-0.0444		9.3278	0.9517	18	-0.0353	
19	0.0425		9.5584	0.9630	19	0.0827	
20	0.0391		9.7567	0.9724	20	-0.0208	
21	-0.0899		10.8141	0.9663	21	-0.1219	
22	-0.0076		10.8217	0.9772	22	-0.0328	
23	-0.0883		11.8698	0.9724	23	-0.1022	
24	0.0276		11.9733	0.9802	24	0.0375	
25	-0.2497		20.5620	0.7168	25	-0.2629	

We can see from the residuals that the residuals are random and there is no auto-correlation left. Thus, suggesting that this model is indeed valid candidate.

ARIMA (2, 0, 0) x (0, 1, 0)12

The model summary and the parameter estimates are given below:

Model: Seasonal ARIMA(2, 0, 0)(0, 1, 0)12

Model Summary

DF	99	Stable	Yes
Sum of Squared Errors	80.4430925	Invertible	Yes
Variance Estimate	0.81255649		
Standard Deviation	0.90141915		
Akaike's 'A' Information Criterion	272.434216		
Schwarz's Bayesian Criterion	280.309135		
RSquare	0.7452899		
RSquare Adj	0.74014424		
MAPE	3.43816534		
MAE	0.70624372		
-2LogLikelihood	266.434216		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	0.9041737	0.0834660	10.83	<.0001*
AR1,2	1	2	-0.5696798	0.0869097	-6.55	<.0001*
Intercept	1	0	-0.0516304	0.1325805	-0.39	0.6978

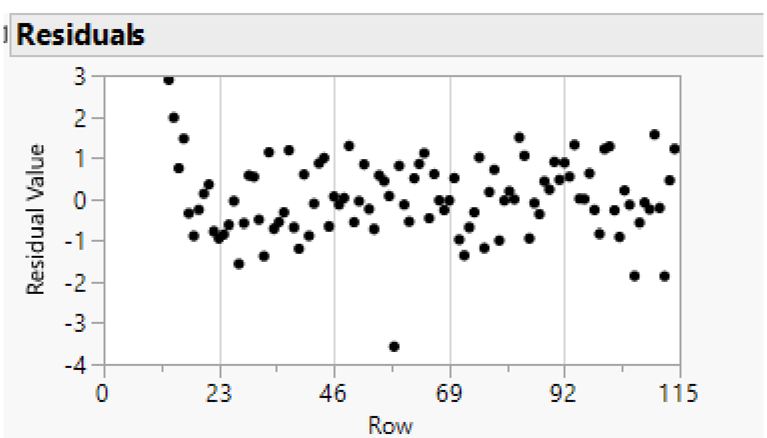
Constant Estimate	Mu
-0.0343603	-0.0516304

As we can see, all the parameters are significant, suggesting this model is valid.

The accuracy measures are given as:

Model	DF	Variance	AIC	SBC	RSquare	MAPE	MAE
Seasonal ARIMA(2, 0, 0)(0, 1, 0)12	99	0.8125565	272.43422	280.30913	0.745	3.438165	0.706244

Residual analysis:



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0949		0.9464	0.3306	1	0.0949	
2	-0.1223		2.5328	0.2818	2	-0.1325	
3	0.1392		4.6096	0.2027	3	0.1694	
4	0.0708		5.1516	0.2721	4	0.0192	
5	-0.0822		5.8902	0.3171	5	-0.0550	
6	-0.0574		6.2545	0.3953	6	-0.0539	
7	0.1018		7.4108	0.3874	7	0.0877	
8	-0.0300		7.5124	0.4825	8	-0.0523	
9	-0.0716		8.0964	0.5245	9	-0.0166	
10	-0.0680		8.6298	0.5676	10	-0.1007	
11	-0.0792		9.3615	0.5886	11	-0.0781	
12	-0.1433		11.7808	0.4634	12	-0.1318	
13	0.0552		12.1446	0.5158	13	0.1108	
14	0.0043		12.1469	0.5945	14	-0.0450	
15	-0.0698		12.7411	0.6223	15	-0.0050	
16	0.0051		12.7443	0.6914	16	-0.0203	
17	-0.0283		12.8439	0.7466	17	-0.0527	
18	-0.0486		13.1422	0.7831	18	-0.0336	
19	0.0404		13.3504	0.8202	19	0.0771	
20	0.0173		13.3889	0.8601	20	-0.0461	
21	-0.1073		14.8965	0.8281	21	-0.1074	
22	-0.0337		15.0473	0.8602	22	-0.0521	
23	-0.0233		15.1203	0.8903	23	-0.0693	
24	-0.0204		15.1768	0.9155	24	0.0063	
25	-0.2604		24.5163	0.4897	25	-0.2637	

We can see from the residuals that the residuals are random and there is no auto-correlation left. Thus, suggesting that this model is indeed valid candidate.

Model selection:

As both the candidate models are valid, we compare the accuracy measures for selecting the best model for Temperature time-series

Model	DF	Variance	AIC	SBC	RSquare	MAPE	MAE
Seasonal ARIMA(2, 0, 1)(0, 1, 0)12	98	0.7926771	270.96838	281.46827	0.753	3.439393	0.705697
Seasonal ARIMA(2, 0, 0)(0, 1, 0)12	99	0.8125565	272.43422	280.30913	0.745	3.438165	0.706244

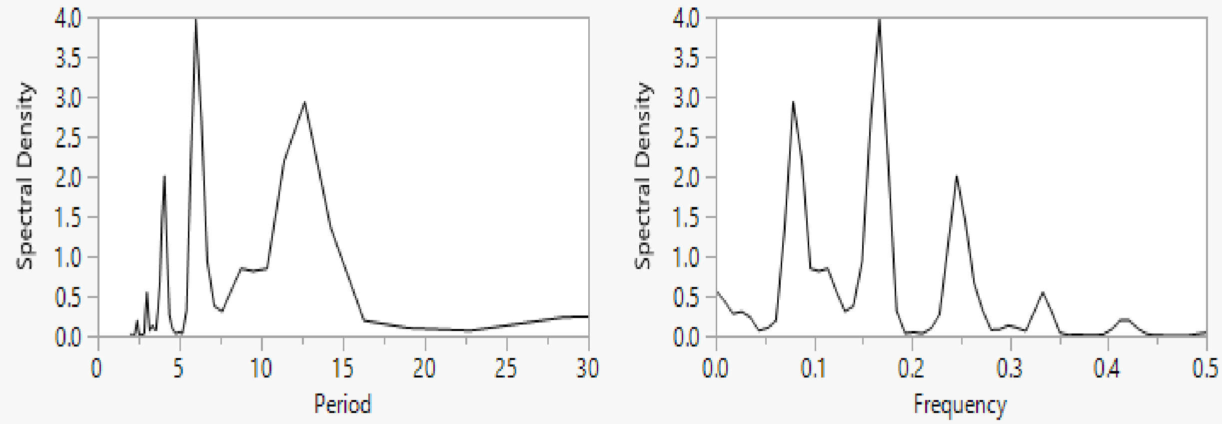
We select ARIMA (2, 0, 0) x $(0, 1, 0)_{12}$ as all the parameters are significant and the accuracy measures are good.

3)

Spectral analysis of Temperature process:

The spectral density of the temperature process is depicted below:

Spectral Density



White Noise test

Fisher's Kappa	11.752839
Prob > Kappa	0.0001323
Bartlett's Kolmogorov-Smirnov	0.3899686

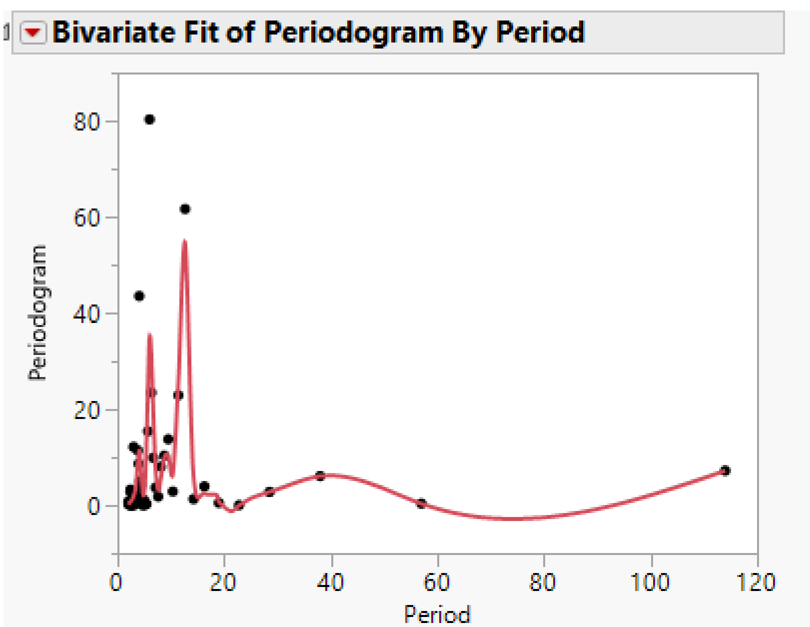
From the above plots, we can see that the Fisher's Kappa statistic is significant as the p-value is below 0.05. Thus, suggesting that there exists one underlying periodic component in it. From the Spectral Density Vs Period plot, we can see that there is a significant and prolonged peak at period 12, which indicates that we have a seasonal period of 12 months.

We save the columns of spectral density and we see the spectral density value of the first spike. We check the corresponding period value and its frequency. That is the period value and the corresponding frequency.

	Period	Frequency	Angular Frequency	Sine	Cosine	Periodogram	Spectral Density
1	•	0	0	0	0	0	0.5737972359
2	114	0.0087719298	0.0551156606	0.2188367252	0.2803771433	7.2105487237	0.4373570825
3	57	0.0175438596	0.1102313212	0.078059956	-0.009363287	0.3523185888	0.2785298681
4	38	0.0263157895	0.1653469818	0.2483136201	-0.212365624	6.0852522984	0.3045790204
5	28.5	0.0350877193	0.2204626424	0.1961964191	-0.101987721	2.786988219	0.2326655386
6	22.8	0.0438596491	0.2755783029	0.0069810432	-0.02407551	0.0358168136	0.0668293335
7	19	0.0526315789	0.3306939635	0.0633784185	-0.069031887	0.5005868422	0.0992107037
8	16.285714286	0.0614035088	0.3858096241	-0.263241661	0.0001666467	3.9498833861	0.1929483797
9	14.25	0.0701754386	0.4409252847	0.1341592044	0.0691254044	1.2982897787	1.3562978157
10	12.666666667	0.0789473684	0.4960409453	-0.542292501	0.887198184	61.628501121	2.9340192783
11	11.4	0.0877192982	0.5511566059	-0.436164199	-0.460376803	22.924602541	2.195479505
12	10.363636364	0.0964912281	0.6062722665	0.0518625048	-0.218680907	2.8791303427	0.8444174746
13	9.5	0.1052631579	0.6613879271	-0.243077651	-0.427030631	13.762188529	0.8108218014
14	8.7692307692	0.1140350877	0.7165035877	0.0787843266	-0.418833894	10.352841634	0.845734289
15	8.1428571429	0.1228070175	0.7716192483	0.1176913761	-0.356735896	8.0433702724	0.5621635682
16	7.6	0.1315789474	0.8267349088	-0.111016228	-0.139883312	1.8178407954	0.3056650636
17	7.125	0.1403508772	0.8818505694	-0.132743616	-0.216874148	3.6853500285	0.3798604821
18	6.7058823529	0.149122807	0.93696623	0.2725763568	0.3154042503	9.9053295474	0.9345758597
19	6.3333333333	0.1578947368	0.9920818906	-0.445200689	-0.462322305	23.480897359	2.7288681354
20	6	0.1666666667	1.0471975512	0.6619776639	0.9851754386	80.300749123	3.9692652824
21	5.7	0.1754385965	1.1023132118	-0.515439835	0.0714486774	15.434638819	2.2173526924
22	5.4285714286	0.1842105263	1.1574288724	-0.044776085	0.0549316844	0.2862761005	0.3207072068
23	5.1818181818	0.1929824561	1.212544533	0.0090737952	0.0436530241	0.1133114558	0.0323067384
24	4.9565217391	0.201754386	1.2676601936	-0.015540589	0.1387442882	1.1110147826	0.0465374383
25	4.75	0.2105263158	1.3227758541	-0.00694362	-0.004467383	0.0038857678	0.028894376
26	4.56	0.2192982456	1.3778915147	0.0719892479	0.0258889933	0.3336034319	0.0942306982
27	4.3846153846	0.2280701754	1.4330071753	-0.180522048	-0.196813694	4.0654588753	0.2689781015
28	4.2222222222	0.2368421053	1.4881228359	-0.116481712	-0.274098762	5.0557928577	1.147169481

The highlighted are the value of a period which suggests 1 period consists of approximately 12 months, and its corresponding frequency = 0.0789 Hz and the first spike spectral density value is spectral density = 2.934

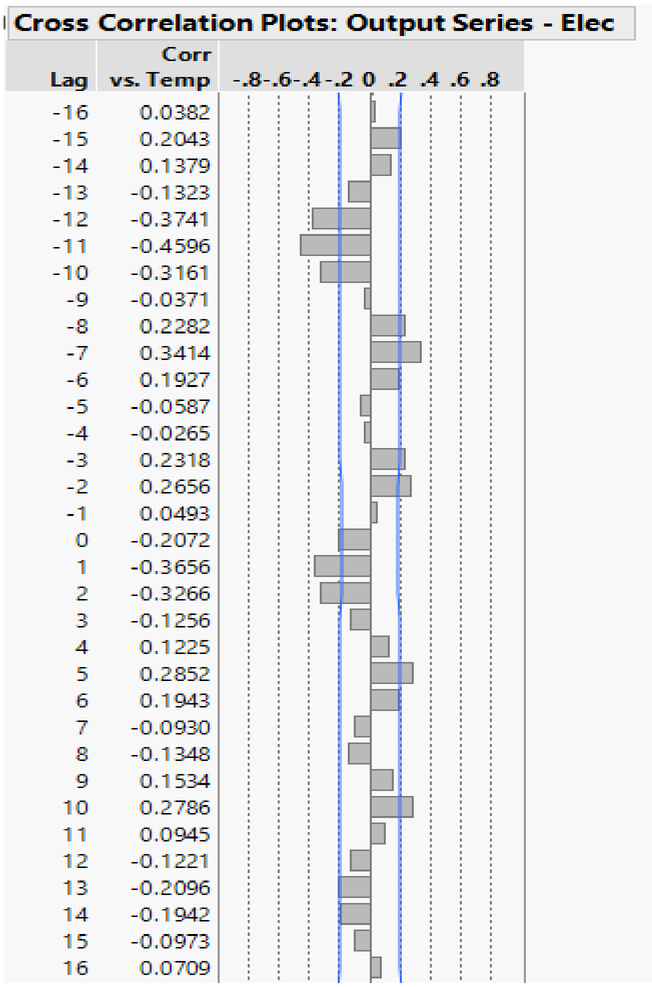
This can be confirmed by the plot below as the highest spike is for approximately 12.



4)

Now we take Temperature as an input and electricity as an output as a cause and effect series. We first see the cross-correlation of the output series to check if pre-whitening is needed.

The cross-correlation plot is as follows:



Cross Correlation Plots: Output Series - Elec

Lag	Corr vs. Temp
-16	0.0382
-15	0.2043
-14	0.1379
-13	-0.1323
-12	-0.3741
-11	-0.4596
-10	-0.3161
-9	-0.0371
-8	0.2282
-7	0.3414
-6	0.1927
-5	-0.0587
-4	-0.0265
-3	0.2318
-2	0.2656
-1	0.0493
0	-0.2072
1	-0.3656
2	-0.3266
3	-0.1256
4	0.1225
5	0.2852
6	0.1943
7	-0.0930
8	-0.1348
9	0.1534
10	0.2786
11	0.0945
12	-0.1221
13	-0.2096
14	-0.1942
15	-0.0973
16	0.0709

As we can see the seasonal components which are significant, and the correlations are significant for negative lags as well, which is unreasonable. Thus, pre-whitening is needed.

We apply pre-whitening using the model ARIMA (2, 0, 0) x $(0, 1, 0)_{12}$ to the input temperature series and obtain the transfer function parameters from the cross-correlation of input series after pre-whitening.

Cross-correlation plot of input series after pre-whitening:

Lag	Prewhiten Corr
-4	0.1297
-3	0.0859
-2	-0.0942
-1	0.1052
0	0.0974
1	-0.0706
2	-0.2188
3	-0.1481
4	0.0731
5	-0.1506
6	-0.1494
7	0.0559
8	-0.1455
9	0.0337
10	0.1865
11	0.0907
12	0.2093
13	0.1651
14	0.1767
15	0.1796
16	0.1995
17	0.1016
18	0.1347
19	0.1147
20	0.0065
21	-0.0563
22	0.0101

From the plot above, we can take b = 2 as the lag starts from lag 2 and r = 0 as there is no exponential decaying pattern and s = 0 as there is w_0 which is of order 0. i.e. ARMAX (2, 0, 0)

Now we put ARMAX (2, 0, 0) and check its residuals to obtain which ARIMA model is adequate for output electricity model.

The parameter estimates and model summary for ARMAX (2, 0, 0) are:

Transfer Function Model (1)

Model Summary

DF	110
Sum of Squared Errors	5384.91081
Variance Estimate	48.9537335
Standard Deviation	6.99669447
Akaike's 'A' Information Criterion	755.60223
Schwarz's Bayesian Criterion	761.039228
RSquare	0.13774369
RSquare Adj	0.13004498
MAPE	4.44594337
MAE	5.20236269
-2LogLikelihood	751.602233

Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temp	Num0,0	0	0	-1.3289	0.355760	-3.74	0.0003*
	Intercept	0	0	148.2911	7.483111	19.82	<.0001*

$$Elec_t = \left(148.2911 - 1.3289 \bullet Temp_{t-2}\right) + e_t$$

We can clearly see that the ARMAX (2, 0, 0) model is valid as all the parameters are significant.

Residual analysis:

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.6356		47.2791	<.0001*	1	0.6356	
2	0.1151		48.8432	<.0001*	2	-0.4847	
3	-0.1003		50.0415	<.0001*	3	0.2176	
4	-0.1625		53.2171	<.0001*	4	-0.2674	
5	-0.2569		61.2237	<.0001*	5	-0.1289	
6	-0.3336		74.8520	<.0001*	6	-0.1381	
7	-0.2407		82.0123	<.0001*	7	0.0818	
8	-0.1315		84.1681	<.0001*	8	-0.2471	
9	-0.1140		85.8046	<.0001*	9	-0.0128	
10	0.0221		85.8668	<.0001*	10	0.2246	
11	0.4093		107.372	<.0001*	11	0.4836	
12	0.6731		166.105	<.0001*	12	0.0732	
13	0.4275		190.033	<.0001*	13	-0.2870	
14	0.0450		190.301	<.0001*	14	0.1197	
15	-0.0944		191.491	<.0001*	15	0.0329	
16	-0.1115		193.168	<.0001*	16	0.0305	
17	-0.2091		199.131	<.0001*	17	-0.1061	
18	-0.3044		211.897	<.0001*	18	-0.0042	
19	-0.2111		218.100	<.0001*	19	0.0557	
20	-0.1077		219.732	<.0001*	20	-0.0880	
21	-0.1020		221.211	<.0001*	21	0.0763	
22	-0.0396		221.436	<.0001*	22	-0.1817	
23	0.2302		229.134	<.0001*	23	0.0914	
24	0.4330		256.683	<.0001*	24	-0.1095	
25	0.2691		267.445	<.0001*	25	0.0719	

From the residuals we can see that there are significant auto-correlations left which should be modeled by an ARIMA model and the seasonal component should be modeled out by a seasonal ARIMA model.

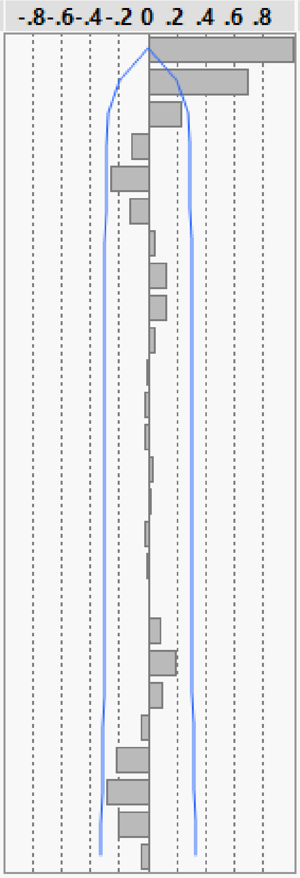
We take seasonal differencing of order 1 and check the residuals ACF and PACF to see which ARIMA model to model the noise for output series

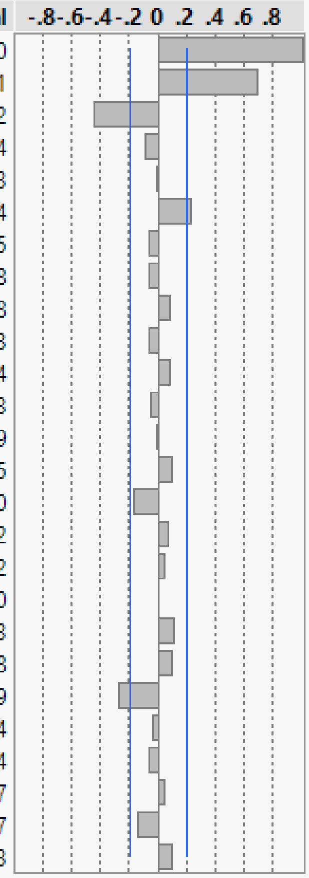
Residual ACF and PACF after seasonal differencing:

Residuals

Row

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.6861	49.4443	<.0001*	1	0.6861
2	0.2352	55.3104	<.0001*	2	-0.4452
3	-0.1110	56.6318	<.0001*	3	-0.0914
4	-0.2530	63.5630	<.0001*	4	-0.0033
5	-0.1275	65.3406	<.0001*	5	0.2294
6	0.0452	65.5665	<.0001*	6	-0.0615
7	0.1190	67.1475	<.0001*	7	-0.0588
8	0.1287	69.0161	<.0001*	8	0.0918
9	0.0443	69.2400	<.0001*	9	-0.0613
10	-0.0094	69.2501	<.0001*	10	0.0824
11	-0.0192	69.2933	<.0001*	11	-0.0423
12	-0.0221	69.3509	<.0001*	12	-0.0079
13	0.0329	69.4801	<.0001*	13	0.0985
14	0.0154	69.5087	<.0001*	14	-0.1580
15	-0.0181	69.5486	<.0001*	15	0.0672
16	-0.0037	69.5503	<.0001*	16	0.0442
17	0.0079	69.5581	<.0001*	17	0.0110
18	0.0827	70.4214	<.0001*	18	0.1113
19	0.1876	74.9185	<.0001*	19	0.1038
20	0.1040	76.3176	<.0001*	20	-0.2669
21	-0.0505	76.6512	<.0001*	21	-0.0394
22	-0.2161	82.8413	<.0001*	22	-0.0574
23	-0.2777	93.1936	<.0001*	23	0.0397
24	-0.2042	98.8626	<.0001*	24	-0.1327
25	-0.0511	99.2227	<.0001*	25	0.0943





From plots above, we can see that there is a cut-off pattern in PACF of order 2 and an exponential decay type pattern in ACF thus we can choose 2 candidate models to model this

1) ARMAX (2, 0, 0) + ARIMA (2, 0, 0) x $(0, 1, 0)_{12}$
2) ARMAX (2, 0, 0) + ARIMA (2, 0, 1) x $(0, 1, 0)_{12}$

ARMAX (2, 0, 0) + ARIMA (2, 0, 0) x (0, 1, 0)12

The model summary and the parameter estimates are given below:

Transfer Function Model (3)

Model Summary

DF	96
Sum of Squared Errors	824.33375
Variance Estimate	8.58681013
Standard Deviation	2.93032594
Akaike's 'A' Information Criterion	503.804355
Schwarz's Bayesian Criterion	514.225036
RSquare	0.62642001
RSquare Adj	0.61498389
MAPE	2.01328613
MAE	2.39681488
-2LogLikelihood	495.804355

Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temp	Num0,0	0	0	-0.8356678	0.3164668	-2.64	0.0097*
Elec	AR1,1	1	1	0.9906212	0.0888796	11.15	<.0001*
Elec	AR1,2	1	2	-0.4449190	0.0885521	-5.02	<.0001*
	Intercept	0	0	0.6018609	0.6303323	0.95	0.3421

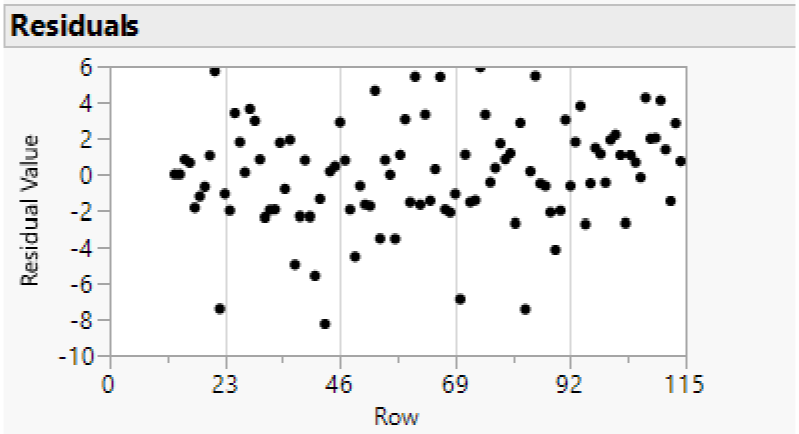
$$\left(1 - B^{12}\right) \cdot Elec_t = \left(0.6019 - 0.8357 \cdot \left(1 - B^{12}\right) \cdot Temp_{t-2}\right) + \left(\frac{1}{\left(\left(1 - 0.9906 \cdot B\right) + 0.4449 \cdot B^2\right)}\right) \cdot e_t$$

As we can see, all the parameters are significant, suggesting this model is valid.

The accuracy measures are given as:

Model	DF	Variance	AIC	SBC	RSquare	MAPE	MAE
Transfer Function Model (3)	96	8.5868101	503.80436	514.22504	0.626	2.013286	2.396815

Residual analysis:



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	-0.0248		0.0645	0.7996	1	-0.0248	
2	0.0451		0.2805	0.8691	2	0.0445	
3	0.0716		0.8299	0.8423	3	0.0740	
4	-0.1855		4.5561	0.3359	4	-0.1854	
5	0.0506		4.8359	0.4362	5	0.0381	
6	0.1276		6.6338	0.3560	6	0.1485	
7	-0.0486		6.8979	0.4396	7	-0.0266	
8	0.1417		9.1632	0.3287	8	0.0884	
9	-0.0687		9.7015	0.3752	9	-0.0668	
10	-0.0338		9.8333	0.4552	10	0.0033	
11	0.0687		10.3828	0.4963	11	0.0422	
12	-0.1572		13.2960	0.3479	12	-0.1323	
13	0.1653		16.5511	0.2206	13	0.1527	
14	-0.0038		16.5529	0.2808	14	-0.0319	
15	-0.0869		17.4742	0.2913	15	-0.0525	
16	0.1104		18.9768	0.2699	16	0.0450	
17	-0.0650		19.5045	0.3004	17	-0.0059	
18	-0.1364		21.8547	0.2385	18	-0.1212	
19	0.2814		31.9748	0.0315*	19	0.2302	
20	-0.0577		32.4049	0.0392*	20	0.0197	
21	0.0182		32.4484	0.0527	21	-0.0565	
22	-0.0990		33.7493	0.0521	22	-0.1892	
23	-0.1158		35.5491	0.0458*	23	0.0231	
24	-0.1211		37.5439	0.0386*	24	-0.1645	
25	0.0332		37.6959	0.0495*	25	0.0552	

There is no significant auto-correlation left and the residual plot is random, thus, suggesting this model is valid.

ARMAX (2, 0, 0) + ARIMA (2, 0, 1) x (0, 1, 0)12

The model summary and the parameter estimates are given below:

Transfer Function Model (4)

Model Summary

DF	95
Sum of Squared Errors	917.908321
Variance Estimate	9.66225135
Standard Deviation	3.10841621
Akaike's 'A' Information Criterion	518.78811
Schwarz's Bayesian Criterion	531.813961
RSquare	0.58401042
RSquare Adj	0.56685621
MAPE	2.12175896
MAE	2.52189329
-2LogLikelihood	508.78811

Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temp	Num0,0	0	0	-0.9472779	0.3112810	-3.04	0.0030*
Elec	AR1,1	1	1	-0.1341455	0.0871759	-1.54	0.1272
Elec	AR1,2	1	2	0.5023514	0.0869433	5.78	<.0001*
Elec	MA1,1	1	1	-0.9999859	0.0316606	-31.58	<.0001*
	Intercept	0	0	0.6494985	0.9424396	0.69	0.4924

$$\left(1 - B^{12}\right) \cdot Elec_t = \left(0.6495 - 0.9473 \cdot \left(1 - B^{12}\right) \cdot Temp_{t-2}\right) + \left(\frac{\left(1 + B\right)}{\left(\left(1 + 0.1341 \cdot B\right) - 0.5024 \cdot B^2\right)}\right) \cdot e_t$$

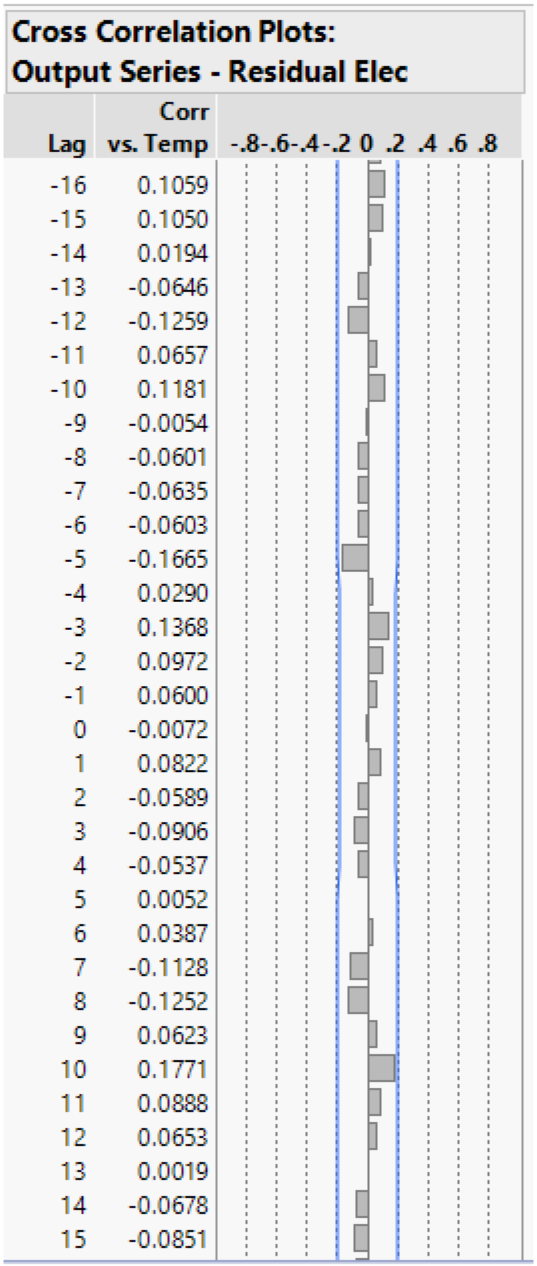
As we can see from the table above, all parameters are not significant for this model.

Thus, we can select ARIMA (2, 0, 0) x (0, 1, 0) to model noise model.

The final model for transfer function is ARMAX (2, 0, 0) + ARIMA (2, 0, 0) x (0, 1, 0)12

To check Independence between the output residual and the input series:

We now take Temperature as input series and Electricity residuals as an output and compute the cross-correlation:



Cross Correlation Plots:
Output Series - Residual Elec

Lag	Corr vs. Temp
-16	0.1059
-15	0.1050
-14	0.0194
-13	-0.0646
-12	-0.1259
-11	0.0657
-10	0.1181
-9	-0.0054
-8	-0.0601
-7	-0.0635
-6	-0.0603
-5	-0.1665
-4	0.0290
-3	0.1368
-2	0.0972
-1	0.0600
0	-0.0072
1	0.0822
2	-0.0589
3	-0.0906
4	-0.0537
5	0.0052
6	0.0387
7	-0.1128
8	-0.1252
9	0.0623
10	0.1771
11	0.0888
12	0.0653
13	0.0019
14	-0.0678
15	-0.0851

From the plot we can see that the cross-correlation is only white noise. Thus, the input series and the residual of the output series are independent.

5)

For the forecast process, we fit optimal seasonal ARIMA models separately to both temperature and electricity.

We already have Optimal ARIMA for the Temperature series.

Electricity:

ACF and PACF of electricity time-series are:

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.7111	59.1764	<.0001*	1	0.7111
2	0.2046	64.1199	<.0001*	2	-0.6090
3	-0.1382	66.3952	<.0001*	3	0.1953
4	-0.2719	75.2841	<.0001*	4	-0.2203
5	-0.2727	84.3053	<.0001*	5	0.0065
6	-0.2485	91.8688	<.0001*	6	-0.2125
7	-0.2559	99.9608	<.0001*	7	-0.1176
8	-0.2825	109.914	<.0001*	8	-0.1917
9	-0.2316	116.668	<.0001*	9	0.0749
10	0.0001	116.668	<.0001*	10	0.2423
11	0.4014	137.355	<.0001*	11	0.4780
12	0.6825	197.744	<.0001*	12	-0.0246
13	0.5513	237.532	<.0001*	13	-0.2932
14	0.1956	242.594	<.0001*	14	0.1599
15	-0.0761	243.366	<.0001*	15	0.0365
16	-0.1990	248.712	<.0001*	16	-0.0469
17	-0.2352	256.256	<.0001*	17	-0.1201
18	-0.2383	264.079	<.0001*	18	0.0078
19	-0.2408	272.152	<.0001*	19	0.0281
20	-0.2645	281.997	<.0001*	20	-0.0303
21	-0.2284	289.415	<.0001*	21	0.1136
22	-0.0612	289.954	<.0001*	22	-0.1322
23	0.2402	298.339	<.0001*	23	0.0645
24	0.4669	330.369	<.0001*	24	-0.1375
25	0.4009	354.245	<.0001*	25	0.0618

As we can see there is some significant seasonal component and AR 2 cut-off pattern. Thus, we will put ARIMA (2, 0, 0) x (0, 1, 0) which is significant by analysis, thus validating our model. We now forecast temperature and electricity separately with their respective ARIMA model.

Forecast for next 6 months

Month	Predicted Temperature	Predicted Electricity
7	23.18910819	121.4144055
8	23.57367519	121.1429699
9	19.58086603	121.9514398
10	15.19227942	127.523679
11	16.99198557	129.3577309
12	19.84934281	129.996468