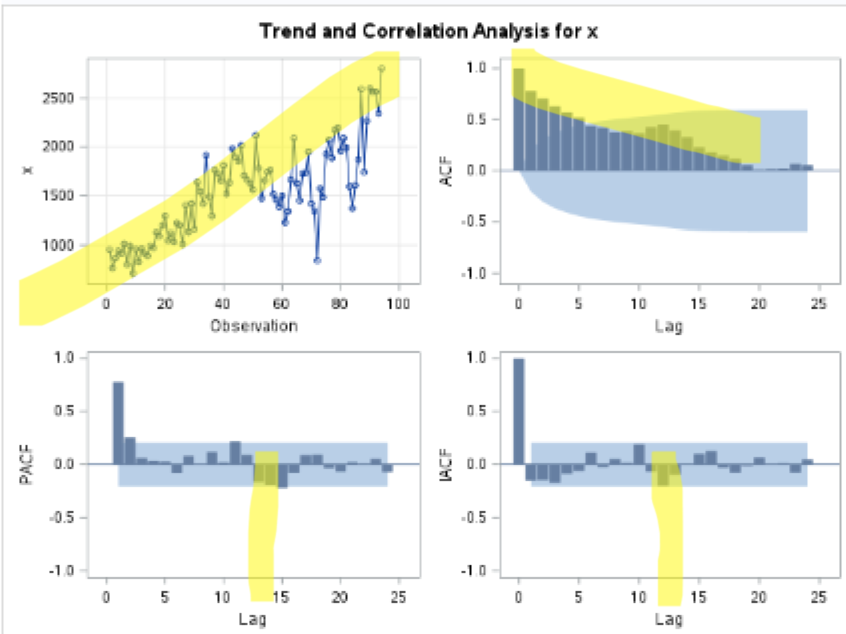
ARIMA Forecasting Variable Temperatures

George Kacoyanis

STA 4853

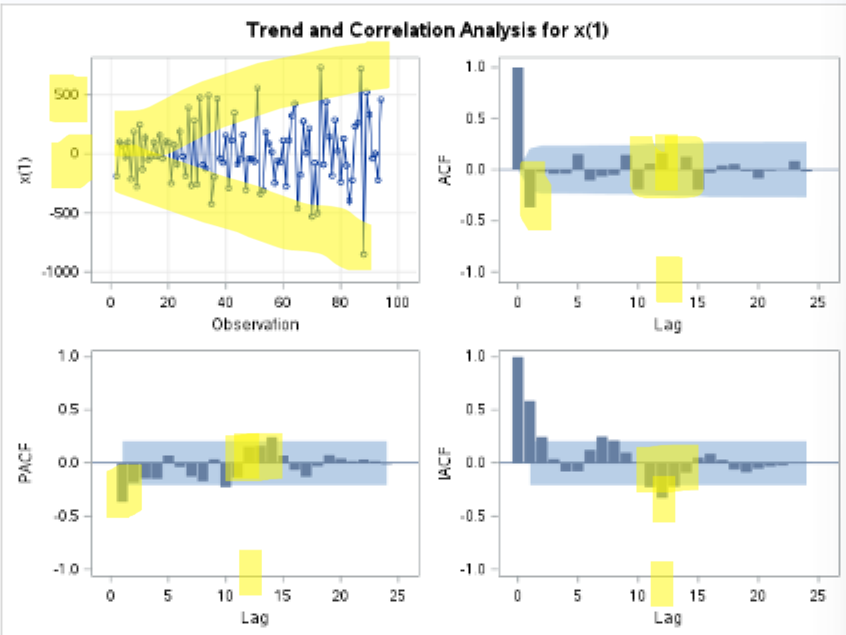
1.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	227.10	6	<.0001	0.778	0.706	0.631	0.571	0.521	0.440
12	332.52	12	<.0001	0.421	0.377	0.386	0.366	0.424	0.451
18	375.72	18	<.0001	0.393	0.328	0.228	0.180	0.147	0.116
24	377.12	24	<.0001	0.055	0.011	0.015	0.020	0.066	0.055



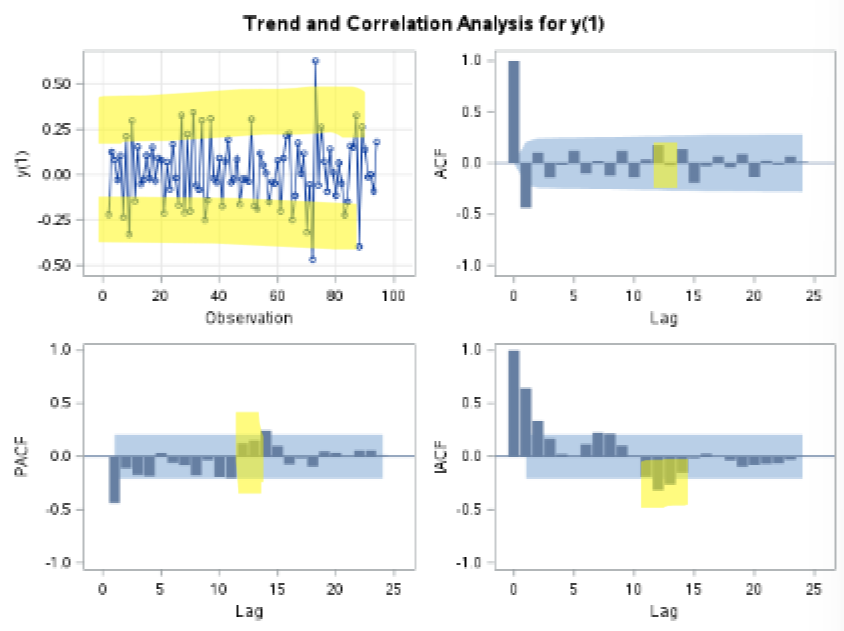
In this series we can see from the ACF that the lags are slowly decreasing to 0 and show no obvious cut offs. Our model also shows that we have an upwards trend. From this we can assume already that our series is not stationary. We start with d=1. We can also see strong positive autocorrelation between our errors in the white noise showing that the lags have strong autocorrelation.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	16.89	6	0.0097	-0.369	-0.024	-0.044	-0.041	0.150	-0.105
12	27.10	12	0.0075	-0.065	-0.053	0.144	-0.193	0.060	0.164
18	34.04	18	0.0125	-0.001	0.128	-0.198	-0.030	0.041	0.055
24	35.98	24	0.0551	-0.015	-0.087	-0.012	0.003	0.086	-0.020



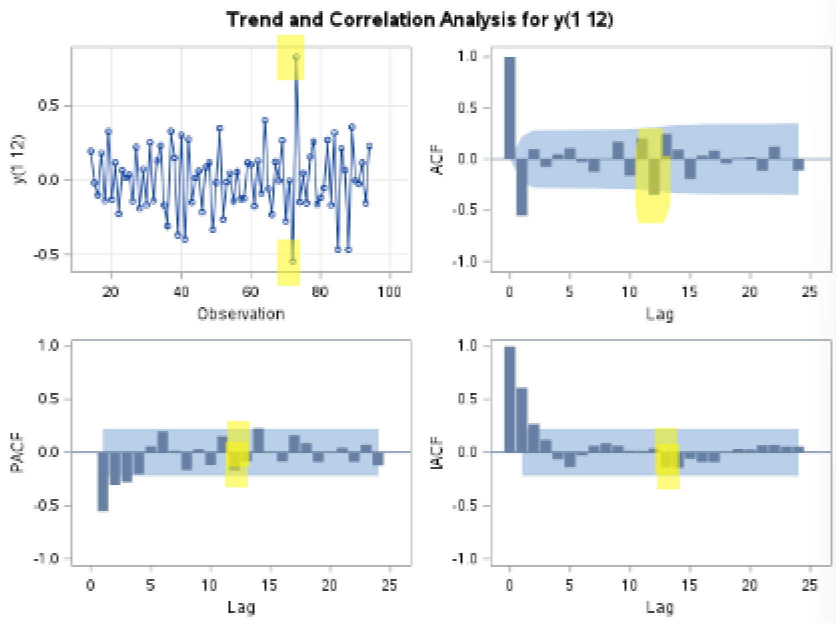
We can see that our ACF has an immediate cutoff to around 0 after the first lag, our series plot no longer has an upward trend either, which can show that our series after differencing we get a stationary model averaging out at around 0. We still see presence of large variance in our series plot, so we can apply a logarithmic transformation to help eliminate some of that variance and make it more normally distributed.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	24.01	6	0.0005	-0.439	0.102	-0.140	-0.017	0.121	-0.101
12	32.73	12	0.0011	0.021	-0.124	0.120	-0.139	0.033	0.176
18	39.87	18	0.0022	-0.022	0.137	-0.192	-0.029	0.065	-0.045
24	43.70	24	0.0083	0.089	-0.136	0.023	-0.015	0.064	0.014



We can see some slight seasonality in the lags where in the IAC and the PACF you can see around lag 12 to have some significant lags. This still leads us to believe that we need seasonal terms. We can assume to use seasonal differencing and the seasonal auto regression since there is autoregression present.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi Sq	Autocorrelations					
6	28.48	6	<.0001	-0.555	0.097	-0.075	0.049	0.106	-0.028
12	51.16	12	<.0001	-0.126	-0.005	0.171	-0.164	0.205	-0.350
18	63.03	18	<.0001	0.247	0.094	-0.195	0.033	0.084	-0.045
24	67.67	24	<.0001	0.012	0.020	-0.112	0.124	-0.006	-0.109



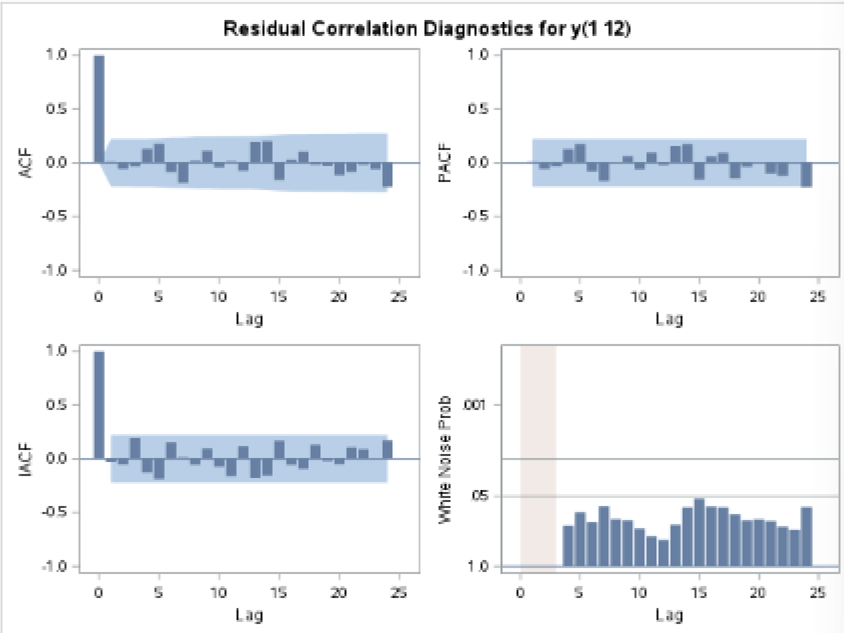
Here we have D=1 and P=1 with AR(1) and MA(1) terms. With this model we can see the Correlation plots all have cut off lags at the beginning. Showing a good fit. We can see the spike in the IACF and PACF at lag 12 disappear and is below the significance.

Constant Estimate	0.002318
Variance Estimate	0.028609
Std Error Estimate	0.169143
AIC	-52.2659
SBC	-42.6881
Number of Residuals	81

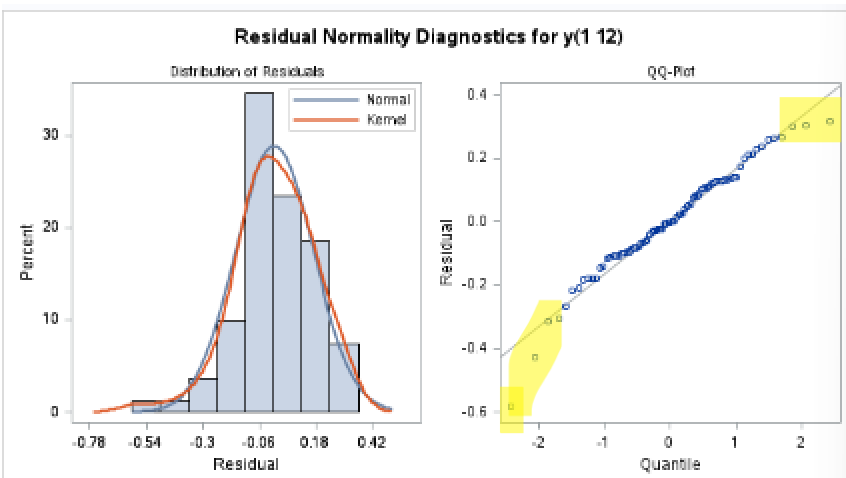
We have good AIC and SBC values currently which are very small.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.59031	0.12242	4.82	<.0001	1
AR1,1	-0.22865	0.13783	-1.66	0.0971	1
AR1,2	-0.28754	0.11289	-2.55	0.0109	12

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.29	3	0.1517	0.010	-0.059	-0.034	0.128	0.177	-0.088
12	10.35	9	0.3232	-0.187	0.016	0.111	-0.042	0.012	-0.074
18	21.93	15	0.1097	0.193	0.199	-0.161	0.028	0.105	-0.020
24	30.85	21	0.0762	-0.030	-0.116	-0.087	-0.022	-0.063	-0.225



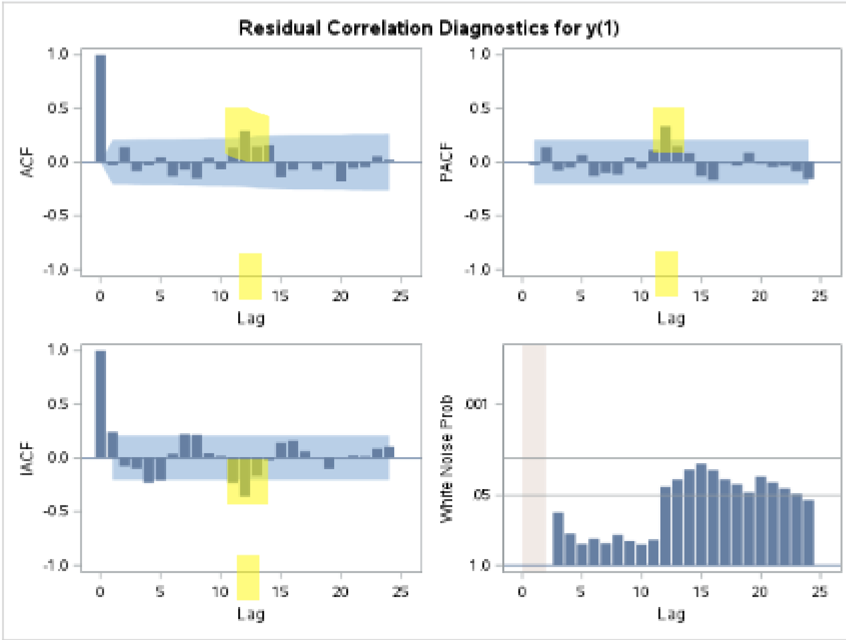
The residual correlation diagnostics plot shows no significant lags and our auto correlation check for residuals shows the p values to be above the significance.



From the histogram of residuals and the Q-Q plot we can see some large outliers, but most of the data fits normal distribution.

From this information, the best model I could find is the ARIMA(1,1,1)(1,1,0)_12.

(b)



You can see that removing the seasonal differencing and seasonal Auto Regression has made the lags around 12 spike and go past the significance on all the autocorrelation plots. This shows that the seasonal pattern is not being accounted for and that all the data is not being expressed by the model.

Constant Estimate	0.009048
Variance Estimate	0.026388
Std Error Estimate	0.162445
AIC	-70.5522
SBC	-62.9544
Number of Residuals	93

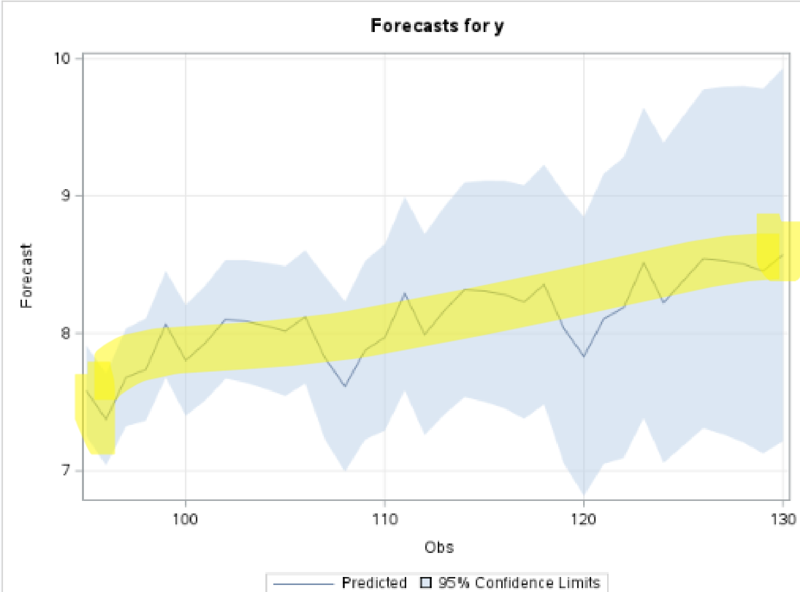
While our AIC and SBC show to be better than with our seasonal terms, we know that the data is not fully expressed by the model from the residual diagnostics.

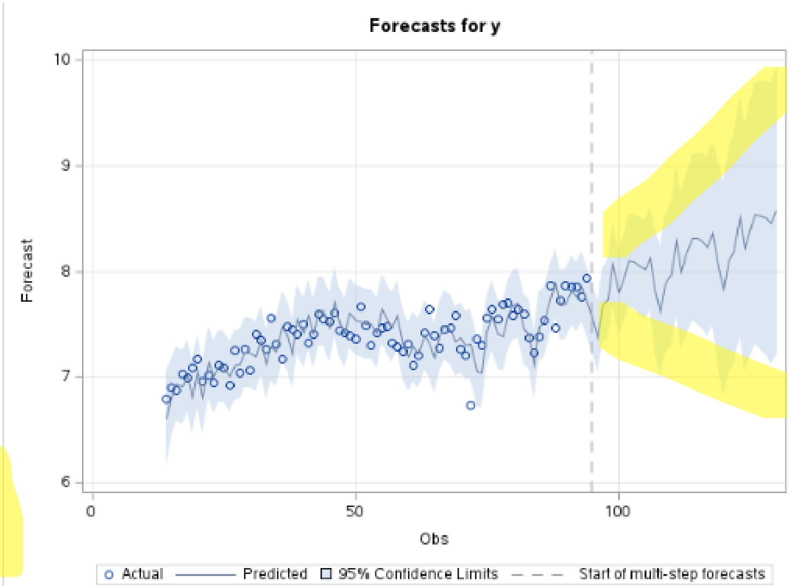
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.01146	0.0049351	2.32	0.0202	0
MA1,1	0.77696	0.10252	7.58	<.0001	1
AR1,1	0.21040	0.15410	1.37	0.1721	1

(c).

Forecasts with no constant of the next 36 months:

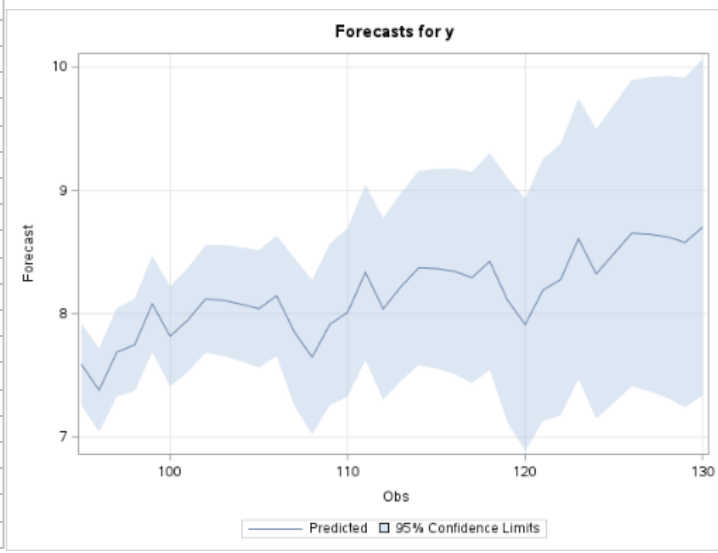
Forecasts for variable y				
Obs	Forecast	Std Error	95% Confidence Limits	
95	7.5844	0.1681	7.2549	7.9140
96	7.3726	0.1709	7.0377	7.7075
97	7.6795	0.1818	7.3233	8.0357
98	7.7367	0.1898	7.3647	8.1088
99	8.0663	0.1980	7.6782	8.4544
100	7.8031	0.2058	7.3998	8.2064
101	7.9327	0.2133	7.5147	8.3508
102	8.1022	0.2205	7.6700	8.5344
103	8.0879	0.2275	7.6420	8.5339
104	8.0555	0.2343	7.5962	8.5149
105	8.0168	0.2410	7.5446	8.4891
106	8.1190	0.2474	7.6341	8.6039
107	7.8199	0.3035	7.2250	8.4148
108	7.6130	0.3163	6.9931	8.2329
109	7.8760	0.3316	7.2261	8.5259
110	7.9705	0.3461	7.2921	8.6489
111	8.2907	0.3600	7.5852	8.9963
112	7.9910	0.3734	7.2593	8.7228
113	8.1673	0.3863	7.4103	8.9244
114	8.3170	0.3987	7.5355	9.0986
115	8.3073	0.4109	7.5020	9.1125
116	8.2834	0.4226	7.4551	9.1117
117	8.2277	0.4341	7.3770	9.0785
118	8.3560	0.4452	7.4834	9.2286
119	8.0350	0.5024	7.0504	9.0196
120	7.8317	0.5184	6.8156	8.8477
121	8.1064	0.5393	7.0494	9.1635
122	8.1876	0.5582	7.0935	9.2817
123	8.5135	0.5768	7.3830	9.6441
124	8.2230	0.5948	7.0573	9.3887
125	8.3838	0.6122	7.1839	9.5837
126	8.5427	0.6291	7.3096	9.7758
127	8.5296	0.6456	7.2641	9.7950
128	8.5040	0.6617	7.2070	9.8010
129	8.4536	0.6774	7.1259	9.7814
130	8.5732	0.6928	7.2154	9.9310



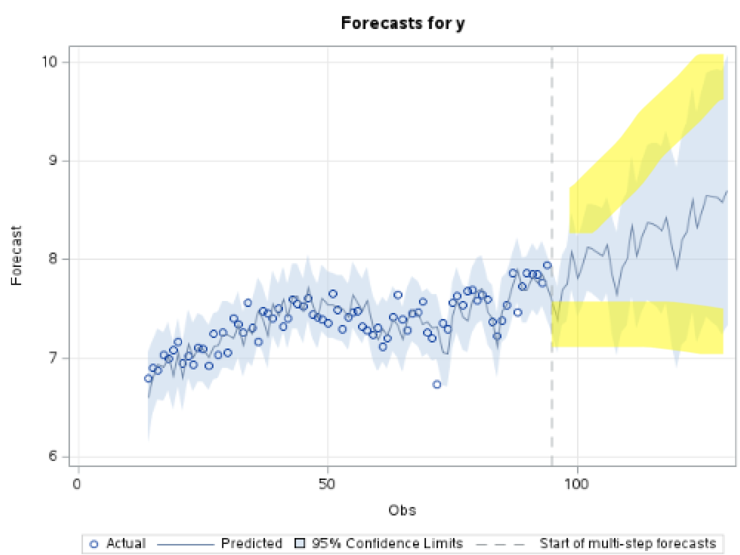


Forecasts with constant of the next 36 months:

Forecasts for variable y				
Obs	Forecast	Std Error	95% Confidence Limits	
95	7.5901	0.1691	7.2586	7.9216
96	7.3793	0.1719	7.0424	7.7162
97	7.6883	0.1828	7.3300	8.0467
98	7.7474	0.1910	7.3731	8.1216
99	8.0788	0.1992	7.6884	8.4693
100	7.8175	0.2070	7.4118	8.2233
101	7.9490	0.2146	7.5284	8.3696
102	8.1204	0.2219	7.6856	8.5553
103	8.1080	0.2289	7.6593	8.5567
104	8.0775	0.2358	7.6154	8.5396
105	8.0407	0.2424	7.5655	8.5159
106	8.1447	0.2489	7.6569	8.6326
107	7.8516	0.3054	7.2531	8.4501
108	7.6476	0.3182	7.0240	8.2713
109	7.9140	0.3336	7.2601	8.5679
110	8.0119	0.3483	7.3293	8.6944
111	8.3354	0.3622	7.6255	9.0453
112	8.0391	0.3757	7.3028	8.7753
113	8.2187	0.3886	7.4569	8.9804
114	8.3717	0.4012	7.5854	9.1581
115	8.3653	0.4134	7.5550	9.1755
116	8.3447	0.4252	7.5113	9.1782
117	8.2924	0.4368	7.4364	9.1484
118	8.4240	0.4480	7.5460	9.3020



119	8.1108	0.5055	7.1201	9.1015
120	7.9118	0.5216	6.8895	8.9341
121	8.1916	0.5427	7.1280	9.2552
122	8.2776	0.5617	7.1766	9.3785
123	8.6084	0.5804	7.4708	9.7460
124	8.3228	0.5985	7.1498	9.4958
125	8.4884	0.6160	7.2811	9.6958
126	8.6523	0.6331	7.4115	9.8930
127	8.6440	0.6497	7.3706	9.9173
128	8.6233	0.6659	7.3182	9.9283
129	8.5778	0.6817	7.2417	9.9138
130	8.7022	0.6971	7.3359	10.0685



Both forecasts have similar plots and constantly increasing standard errors. But we can see that when we include the constant, our confidence intervals and our forecasts rise and we get greater values for our forecast showing a slightly greater upward trend.

Forecasts With Constant:

126	8.6523	0.6331	7.4115	9.8930
127	8.6440	0.6497	7.3706	9.9173
128	8.6233	0.6659	7.3182	9.9283
129	8.5778	0.6817	7.2417	9.9138
130	8.7022	0.6971	7.3359	10.0685

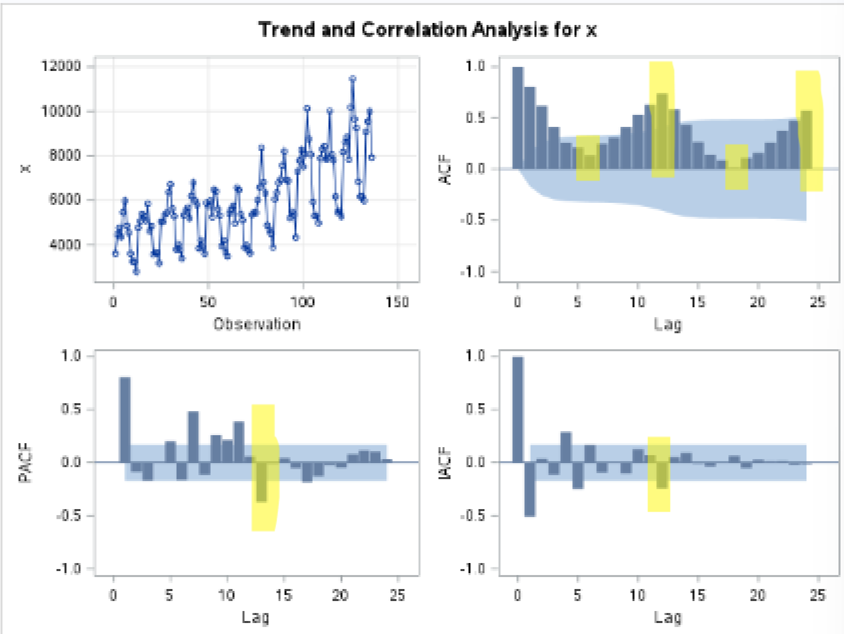
Forecasts Without Constant:

126	8.5427	0.6291	7.3096	9.7758
127	8.5296	0.6456	7.2641	9.7950
128	8.5040	0.6617	7.2070	9.8010
129	8.4536	0.6774	7.1259	9.7814
130	8.5732	0.6928	7.2154	9.9310

2.

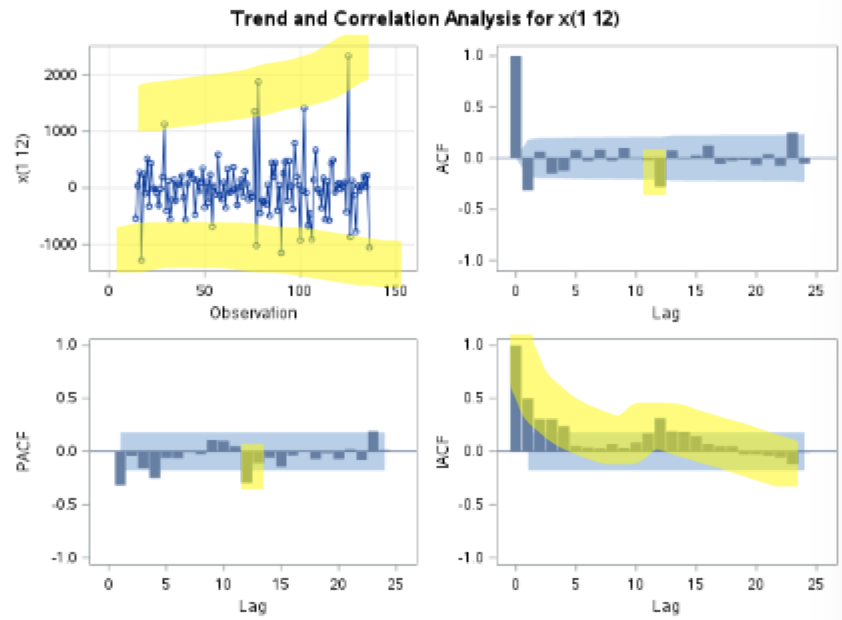
Immediately after running an ARIMA process on the data in BUS.txt we can see the immediate need for differencing and seasonal differencing with the constant acf, obvious patterns in the data series, and lag spikes at lag 12 for all the acf plots.

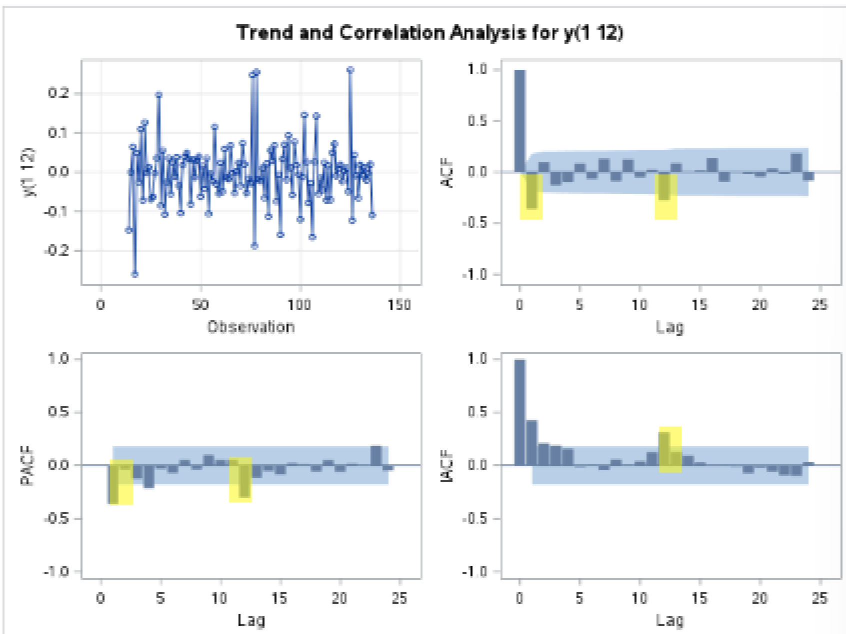
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	185.83	6	<.0001	0.805	0.618	0.411	0.257	0.211	0.135
12	417.41	12	<.0001	0.249	0.306	0.411	0.531	0.630	0.738
18	513.22	18	<.0001	0.583	0.434	0.263	0.139	0.086	0.013
24	645.34	24	<.0001	0.107	0.158	0.258	0.374	0.475	0.572



Looking at the trend and correlation analysis there is a much better acf and pacf. Our IACF shows that there may be some slight over differencing from the constant decrease in the significance of the lags. The series plot has large variance between the values and outliers, which may suggest that a transformation might be needed. The y=log(x) transformation might help with these issues.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	19.04	6	0.0041	-0.318	0.063	-0.150	-0.126	0.079	-0.027
12	32.58	12	0.0011	0.085	-0.026	0.099	-0.002	-0.023	-0.283
18	36.28	18	0.0065	0.074	-0.008	0.026	0.125	-0.057	-0.026
24	48.09	24	0.0025	-0.017	-0.065	0.038	-0.074	0.250	-0.054





We get about the same graph ACF and PACF, but our series plot shows slightly less variance and helps to make the numbers we are working with slightly smaller. Also the IACF shows to have less evidence of over differencing slightly. From our ACF we can see that there may be evidence of an AR(1) and MA(1) process for seasonal and non-seasonal terms (p=1,q=1,P=1,Q=1).

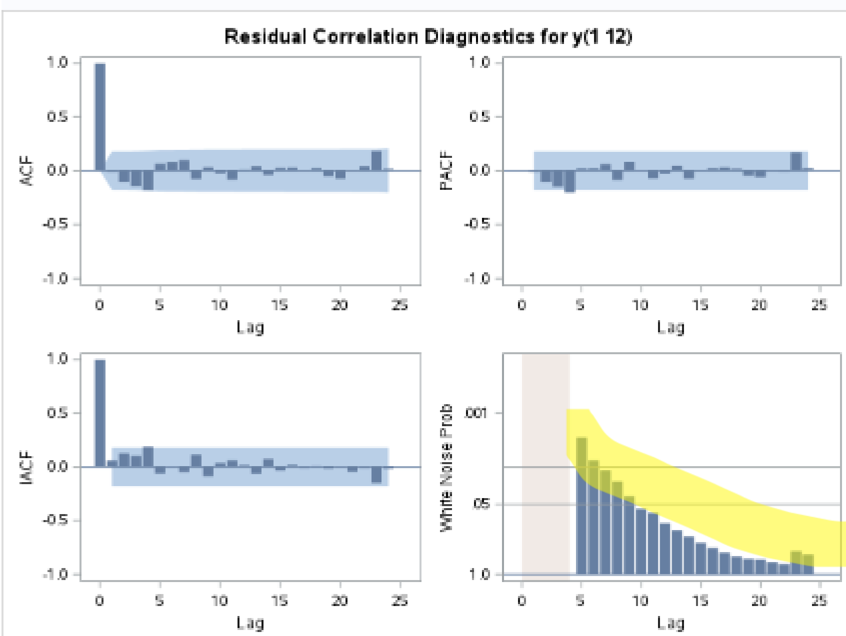
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.0007189	0.0012061	-0.60	0.5511	0
MA1,1	0.01150	0.10664	0.11	0.9142	1
MA1,2	0.98849	1.73198	0.57	0.5682	12
AR1,1	-0.39926	0.10005	-3.99	<.0001	1
AR1,2	0.26328	0.12005	2.19	0.0283	12

Constant Estimate	-0.00082
Variance Estimate	0.003594
Std Error Estimate	0.059949
AIC	-316.009
SBC	-301.948
Number of Residuals	123

The AIC and SBC are very small which is favorable in a model, but our maximum likelihood estimates show that we might not need a non-seasonal AR(1) term and that there is strong evidence suggesting a seasonal moving average term is necessary. Removing the non-seasonal AR(1) term might be favorable.

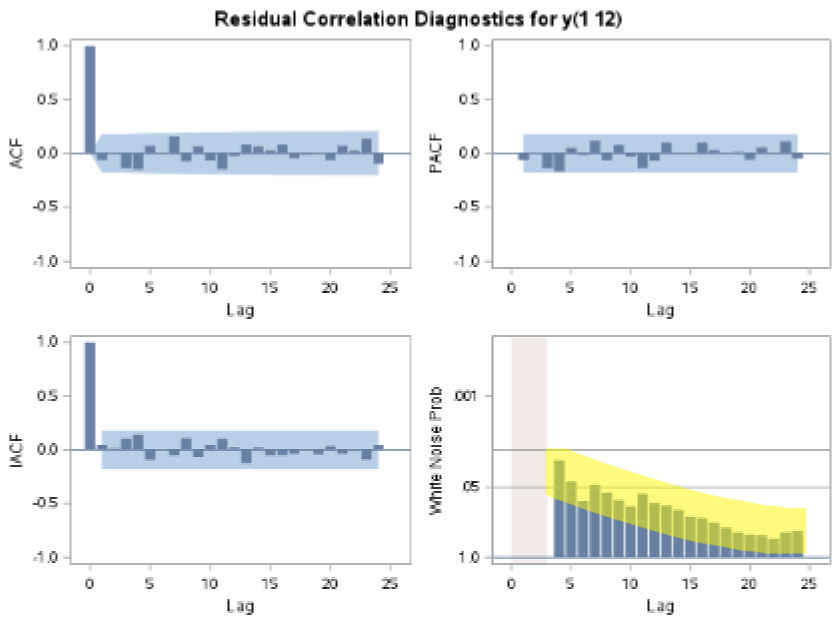
With both AR(1) terms:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.75	2	0.0076	-0.011	-0.106	-0.144	-0.182	0.064	0.080
12	12.99	8	0.1122	0.097	-0.078	0.032	-0.026	-0.083	0.010
18	13.73	14	0.4700	0.041	-0.040	0.025	0.029	0.002	0.022
24	20.51	20	0.4264	-0.053	-0.074	0.002	0.042	0.185	0.016



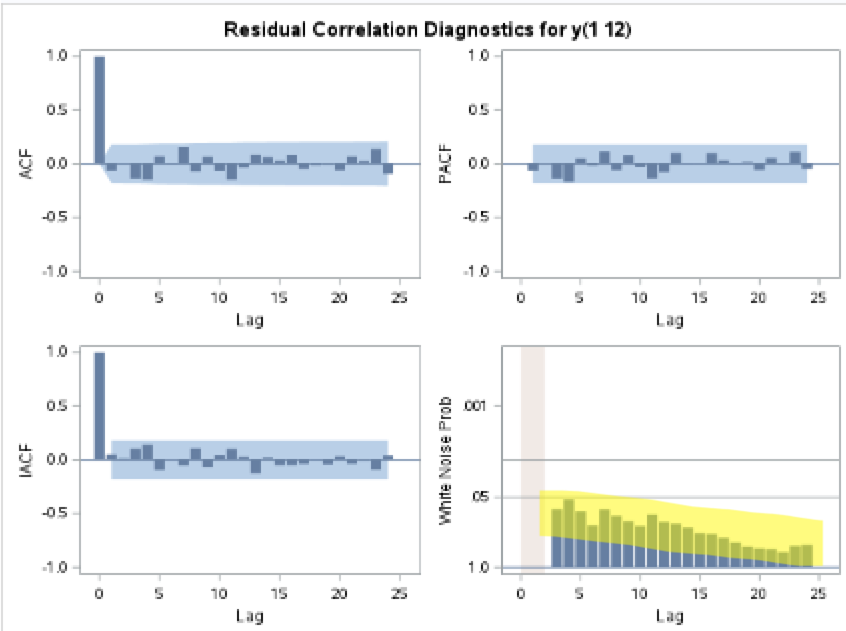
With only the seasonal AR(1) term:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.47	3	0.0910	-0.062	0.004	-0.138	-0.149	0.071	-0.006
12	14.68	9	0.1000	0.156	-0.074	0.065	-0.067	-0.147	-0.025
18	17.55	15	0.2871	0.082	0.063	0.027	0.082	-0.045	-0.013
24	23.35	21	0.3258	-0.006	-0.063	0.072	0.024	0.136	-0.097



With no AR(1) terms:

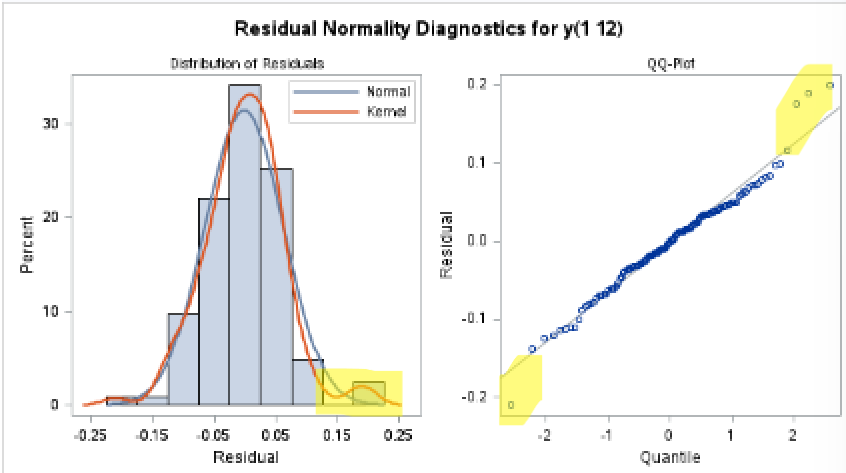
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.47	4	0.1668	-0.065	0.004	-0.139	-0.148	0.070	-0.006
12	14.65	10	0.1452	0.155	-0.071	0.067	-0.066	-0.147	-0.033
18	17.57	16	0.3495	0.083	0.062	0.027	0.083	-0.045	-0.014
24	23.31	22	0.3847	-0.007	-0.062	0.070	0.024	0.138	-0.095

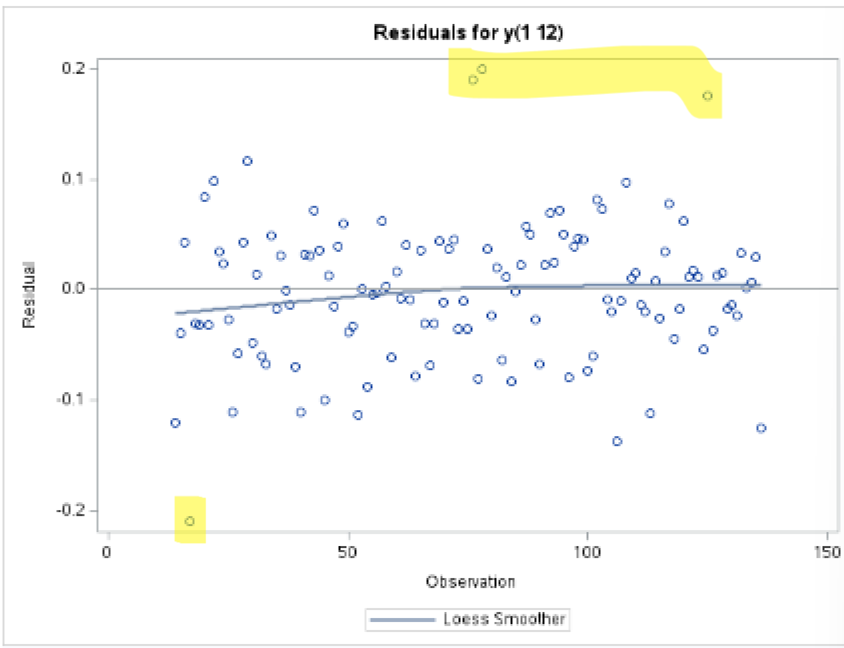


When the AR(1) terms are removed, the white noise probability plot improves and shows that the lags are not significantly correlated with each other. We also get improvements in our AIC and SBC which show the model to be favorable as well.

Constant Estimate	-0.00052
Variance Estimate	0.004091
Std Error Estimate	0.063962
AIC	-316.979
SBC	-308.542
Number of Residuals	123

Checking the Residual Normality Diagnostics, it shows the residuals to be distributed normally with some outliers. The residual series plot also shows that while there are some outliers most of the residuals are distributed normally.

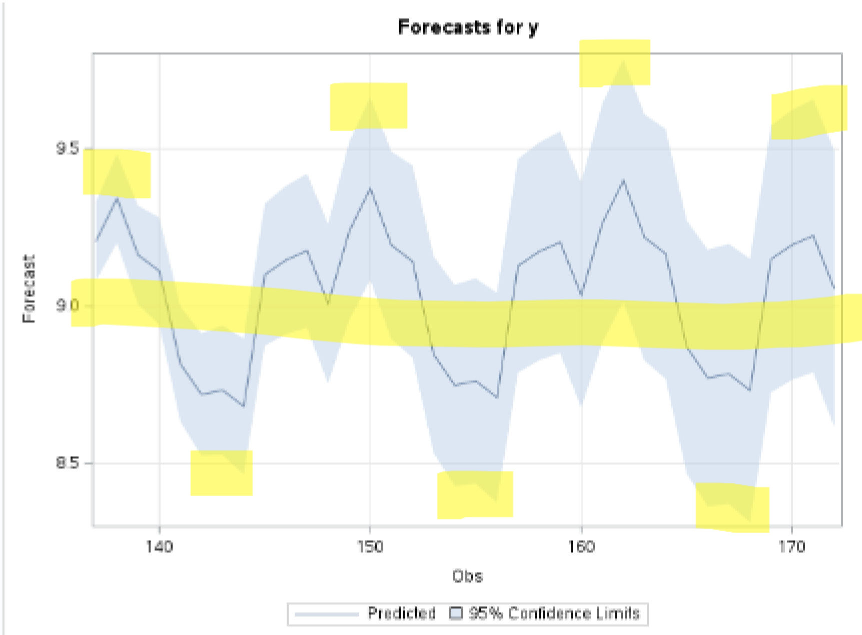


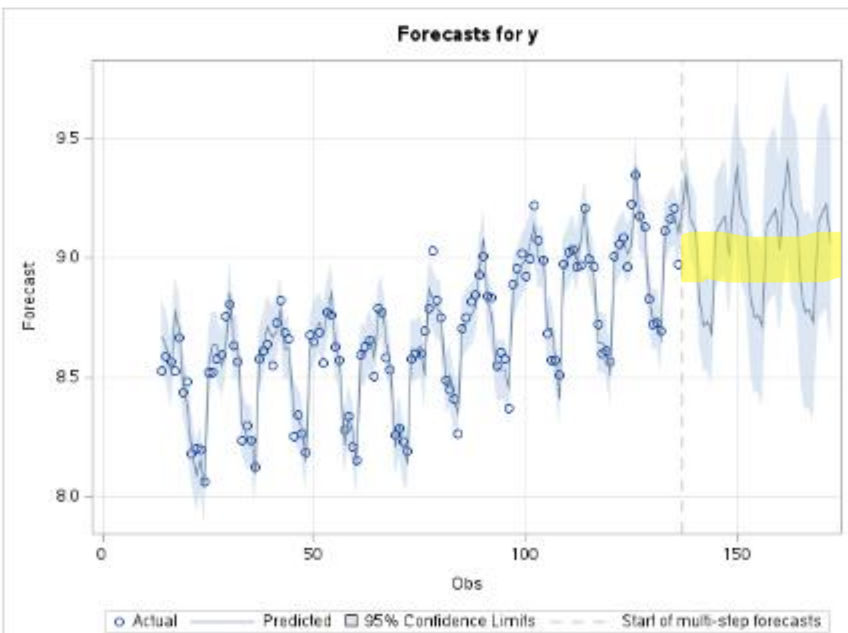


Here we can see that we have our best fit model ARIMA(0,1,1)(0,1,1)_12

(b).

Obs	osforecast	osL95	osU95
101	8588.30	7569.27	9744.53
102	9327.68	8221.08	10583.23
103	8121.44	7158.19	9214.30
104	8119.91	7157.11	9212.23
105	6029.67	5314.88	6840.60
106	6054.87	5337.20	6869.05
107	5337.05	4704.51	6054.64
108	4495.82	3963.01	5100.28
109	7798.70	6874.46	8847.21
110	8166.77	7198.90	9264.75
111	8539.88	7527.81	9688.01
112	7994.64	7047.20	9069.45
113	8796.62	7754.14	9979.25
114	9926.74	8750.38	11261.24
115	8274.03	7293.65	9386.20
116	7527.19	6635.46	8538.76
117	5682.12	5009.10	6445.57
118	5688.18	5014.55	6452.31
119	5592.00	4929.83	6343.11
120	4934.47	4350.20	5597.22
121	8062.46	7107.83	9145.30
122	8460.08	7458.38	9596.32
123	8729.86	7696.21	9902.32
124	8251.44	7274.45	9359.64
125	8527.73	7518.03	9673.04
126	11896.07	10487.56	13493.75
127	9524.95	8397.23	10804.11
128	9108.48	8030.18	10331.58
129	6950.92	6128.15	7884.15
130	6234.92	5497.00	7071.89
131	6312.07	5565.11	7159.29
132	5764.60	5082.48	6538.27
133	9056.24	7984.67	10271.62
134	9478.07	8356.60	10750.05
135	9702.46	8554.44	11004.55

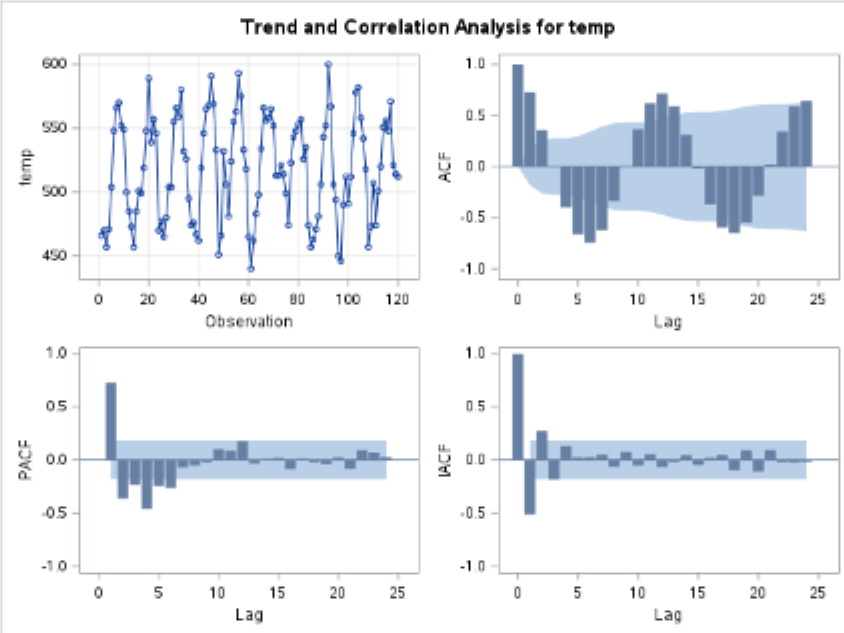


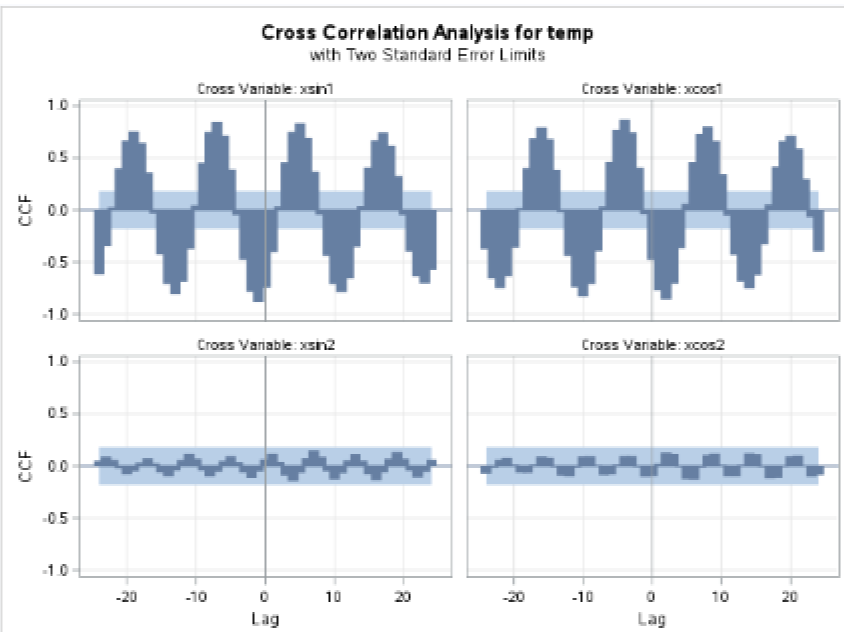


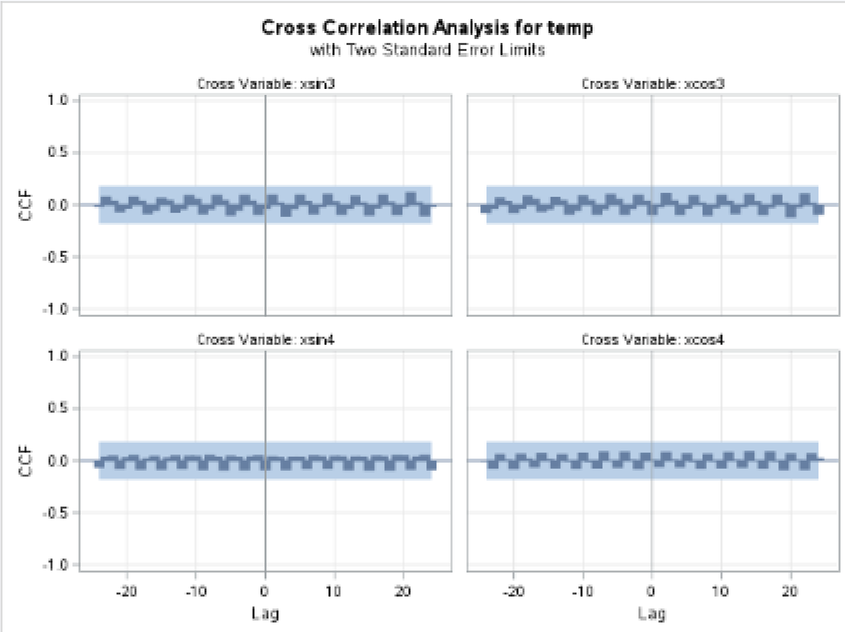
From the forecasts we can see that the repeating pattern continues and our forecasted values somewhat plateau. We have very large spreads in our confidences intervals and the revolving up and down patter also effects the spread of our confidence intervals to be lower or higher depending on what part of the forecasted pattern you are looking at.

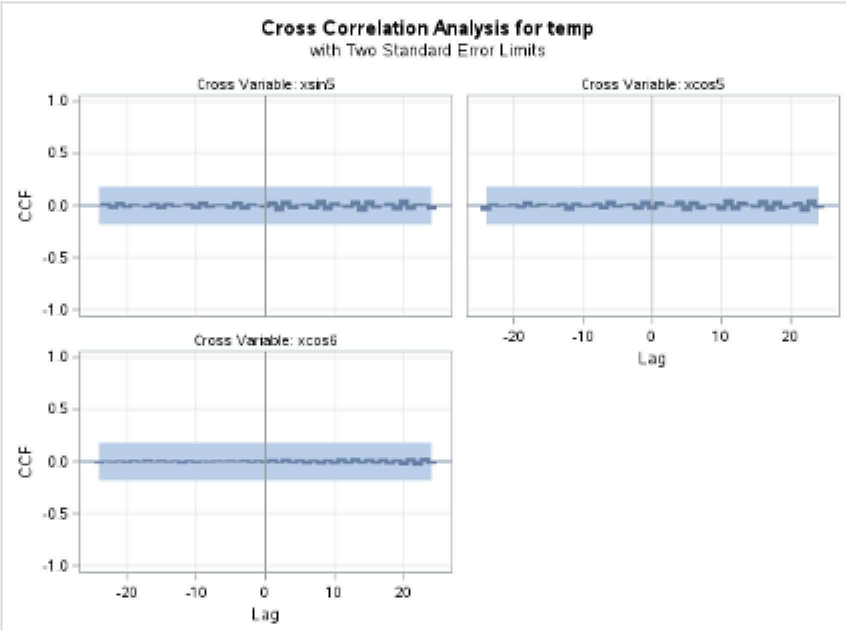
3.

(a)


Trend and Correlation Analysis for temp
temp
Observation
ACF
PACF
IACF
Lag


Cross Correlation Analysis for temp
with Two Standard Error Limits
Cross Variable: xsin1
Cross Variable: xcos1
Cross Variable: xsin2
Cross Variable: xcos2
CCF
Lag


Cross Correlation Analysis for temp
with Two Standard Error Limits
Cross Variable: xsin3
Cross Variable: xcos3
Cross Variable: xsin4
Cross Variable: xcos4
CCF
Lag



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.10833	1.61229	321.35	<.0001	0	temp	0
NUM1	-41.96663	2.28012	-18.41	<.0001	0	xsin1	0
NUM2	-26.98699	2.28012	-11.84	<.0001	0	xcos1	0
NUM3	3.66617	2.28012	1.61	0.1079	0	xsin2	0
NUM4	-5.53333	2.28012	-2.43	0.0152	0	xcos2	0
NUM5	-2.26667	2.28012	-0.99	0.3202	0	xsin3	0
NUM6	-5.65000	2.28012	-2.48	0.0132	0	xcos3	0
NUM7	-5.13842	2.28012	-2.25	0.0242	0	xsin4	0
NUM8	0.21667	2.28012	0.10	0.9243	0	xcos4	0
NUM9	-1.35004	2.28012	-0.59	0.5538	0	xsin5	0
NUM10	-2.36301	2.28012	-1.04	0.3000	0	xcos5	0
NUM11	-0.59167	1.61229	-0.37	0.7136	0	xcos6	0

Constant Estimate	518.1083
Variance Estimate	311.938
Std Error Estimate	17.66177
AIC	1041.039
SBC	1074.488
Number of Residuals	120

Here it shows that our slope estimates for xcos6 and xcos4 are small and might be not important in the ARIMA process. We start by excluding xcos6 because it's standard error stands our and the pvalue is very high.

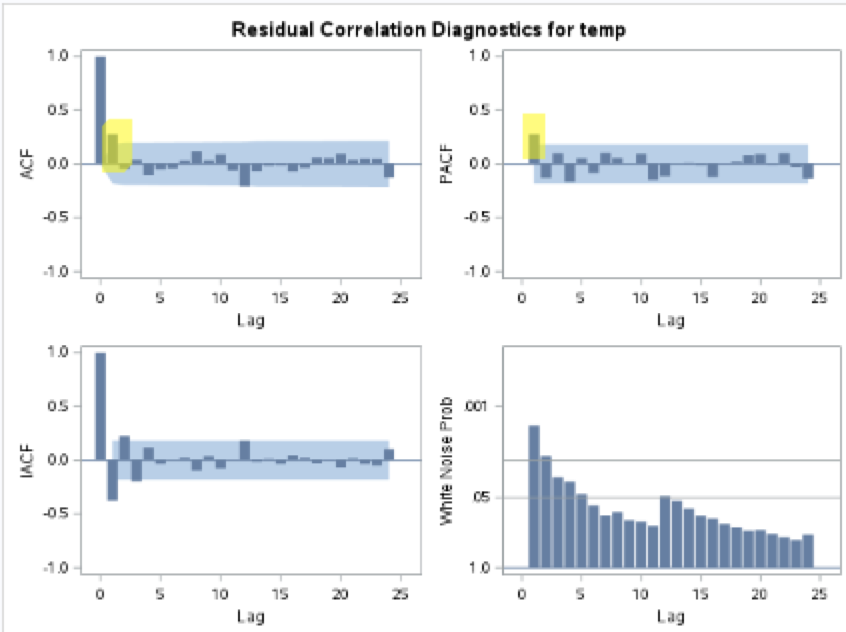
Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.10833	1.60115	323.58	<.0001	0	temp	0
NUM1	-41.96663	2.26437	-18.53	<.0001	0	xsin1	0
NUM2	-26.98699	2.26437	-11.92	<.0001	0	xcos1	0
NUM3	3.66617	2.26437	1.62	0.1054	0	xsin2	0
NUM4	-5.53333	2.26437	-2.44	0.0145	0	xcos2	0
NUM5	-2.26667	2.26437	-1.00	0.3168	0	xsin3	0
NUM6	-5.65000	2.26437	-2.50	0.0126	0	xcos3	0
NUM7	-5.13842	2.26437	-2.27	0.0233	0	xsin4	0
NUM8	0.21667	2.26437	0.10	0.9238	0	xcos4	0
NUM9	-2.36301	2.26437	-1.04	0.2967	0	xcos5	0

Here we see xcos4 to be low and potentially insignificant. We can remove that one and then look at the maximum likelihood extimates again.

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.10833	1.59399	325.04	<.0001	0	temp	0
NUM1	-41.96663	2.25424	-18.62	<.0001	0	xsin1	0
NUM2	-26.98699	2.25424	-11.97	<.0001	0	xcos1	0
NUM3	3.66617	2.25424	1.63	0.1039	0	xsin2	0
NUM4	-5.53333	2.25424	-2.45	0.0141	0	xcos2	0
NUM5	-2.26667	2.25424	-1.01	0.3147	0	xsin3	0
NUM6	-5.65000	2.25424	-2.51	0.0122	0	xcos3	0
NUM7	-5.13842	2.25424	-2.28	0.0226	0	xsin4	0
NUM8	-2.36301	2.25424	-1.05	0.2945	0	xcos5	0

Here we see that all estimates are high, but xsin2, xsin3, and xcos5 have very high pvalues well above the significance. We start with xsin3 and xcos5 one at a time because they have the highest pvalues.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	11.68	6	0.0695	0.275	-0.046	0.039	-0.104	-0.051	-0.042
12	21.29	12	0.0463	0.032	0.118	0.033	0.088	-0.062	-0.210
18	23.35	18	0.1775	-0.068	-0.019	-0.013	-0.070	-0.037	0.057
24	28.42	24	0.2429	0.053	0.093	0.038	0.048	0.047	-0.128



We see slight evidence of AR(1) process from the cutoff at lag 1 and there could be a case for a MA(1) process

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.10833	1.59407	325.02	<.0001	0	temp	0
NUM1	-41.96663	2.25435	-18.62	<.0001	0	xsin1	0
NUM2	-26.98699	2.25435	-11.97	<.0001	0	xcos1	0
NUM3	3.66617	2.25435	1.63	0.1039	0	xsin2	0
NUM4	-5.53333	2.25435	-2.45	0.0141	0	xcos2	0
NUM5	-5.65000	2.25435	-2.51	0.0122	0	xcos3	0
NUM6	-5.13842	2.25435	-2.28	0.0226	0	xsin4	0
NUM7	-2.36301	2.25435	-1.05	0.2945	0	xcos5	0

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.10833	1.59476	324.88	<.0001	0	temp	0
NUM1	-41.96663	2.25534	-18.61	<.0001	0	xsin1	0
NUM2	-26.98699	2.25534	-11.97	<.0001	0	xcos1	0
NUM3	3.66617	2.25534	1.63	0.1040	0	xsin2	0
NUM4	-5.53333	2.25534	-2.45	0.0141	0	xcos2	0
NUM5	-5.65000	2.25534	-2.51	0.0122	0	xcos3	0
NUM6	-5.13842	2.25534	-2.28	0.0227	0	xsin4	0

Here we see that our pvalue is not very above the significance but enough to take it out and check the AIC and SBC for the fit.

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.10833	1.60621	322.57	<.0001	0	temp	0
NUM1	-41.96663	2.27153	-18.48	<.0001	0	xsin1	0
NUM2	-26.98699	2.27153	-11.88	<.0001	0	xcos1	0
NUM3	-5.53333	2.27153	-2.44	0.0149	0	xcos2	0
NUM4	-5.65000	2.27153	-2.49	0.0129	0	xcos3	0
NUM5	-5.13842	2.27153	-2.26	0.0237	0	xsin4	0

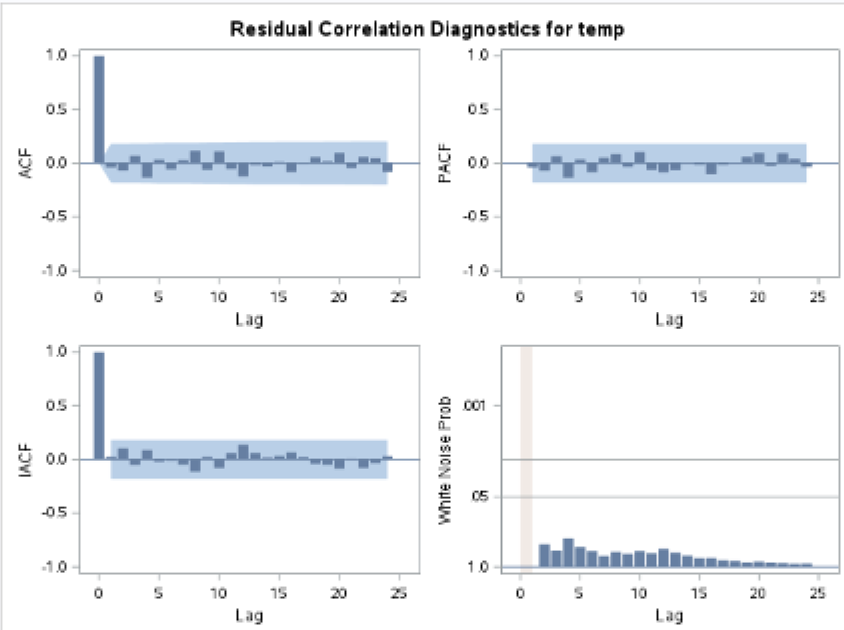
Constant Estimate	518.1083
Variance Estimate	309.5897
Std Error Estimate	17.59516
AIC	1034.62
SBC	1051.345
Number of Residuals	120

Here we see that after removing the sines and cosines we can see that these cosines and sines accurately fit the ARIMA model and see improvements in the AIC and SBC.

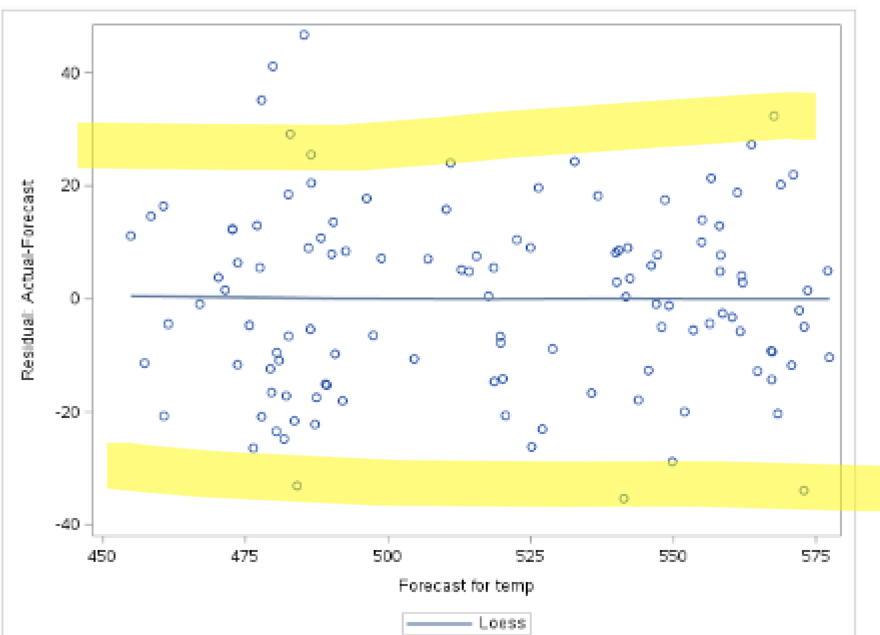
We see that our PACF in our residual correlation diagnostics for temp that we have a cut off at lag 1 from above the significance, so we apply an MA(1) term to the model.

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	518.19807	2.09309	247.58	<.0001	0	temp	0
MA1,1	-0.36548	0.08884	-4.11	<.0001	1	temp	0
NUM1	-41.87785	2.88710	-14.51	<.0001	0	xsin1	0
NUM2	-26.83130	2.87610	-9.33	<.0001	0	xcos1	0
NUM3	-5.44357	2.66062	-2.05	0.0408	0	xcos2	0
NUM4	-5.64730	2.31676	-2.44	0.0148	0	xcos3	0
NUM5	-4.98457	1.90206	-2.62	0.0088	0	xsin4	0

When applying the MA(1) term we see that the p-value Is almost 0 showing there to be strong evidence of a moving average.



Here we see that there is little autocorrelation between the lags and there are no lags past the significance. This helps to show that we have a good fit to

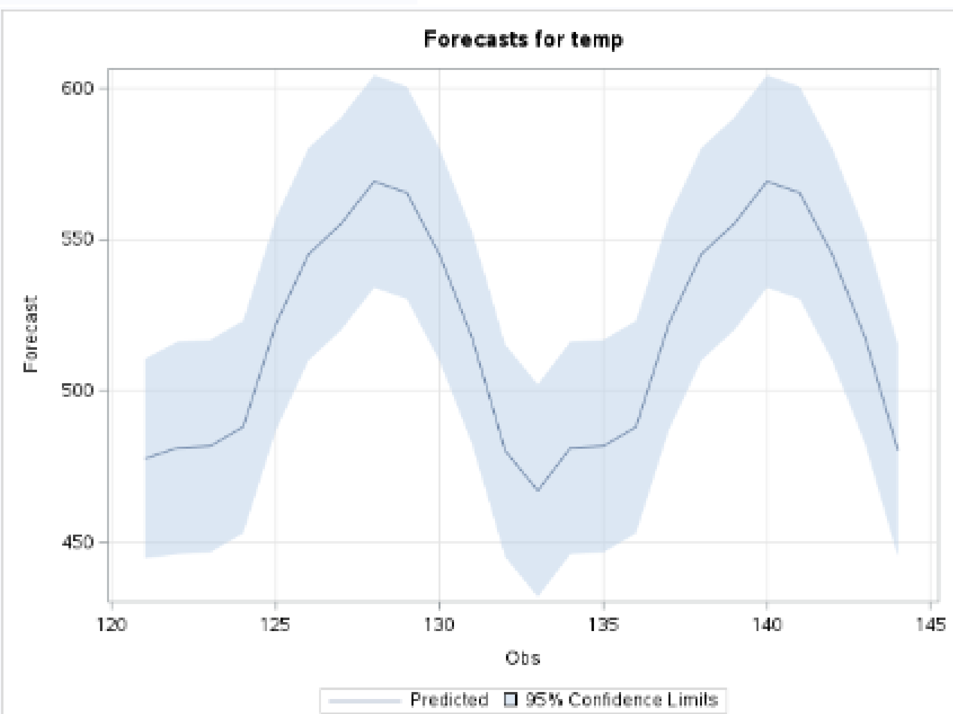


With our residuals forecast we can see that the residual variance remains constant. You can also see this in the Maximum likelihood plots when the STD deviation plateaus to 1 value.

Xt= 518.19807 - 41.87785sin((2*pi*t)/12) – 26.83130cos((2*pi*t)/12) - 5.44357cos((4*pi*t)/12) - 5.64730cos(6*pi*t)/12) -4.9857sin((8*pi*t)/12) + et – 0.36548et-1

Nonseasonal stationary ARMA process and trend consisting of Cosines and sines

Forecasts for variable temp				
Obs	Forecast	Std Error	95% Confidence Limits	
121	477.6385	16.8378	444.6370	510.6399
122	481.2010	17.9271	446.0646	516.3374
123	481.7638	17.9271	446.6274	516.9002
124	488.1042	17.9271	452.9677	523.2406
125	522.0907	17.9271	486.9543	557.2272
126	545.2331	17.9271	510.0967	580.3696
127	555.3350	17.9271	520.1986	590.4715
128	569.2723	17.9271	534.1358	604.4087
129	565.5195	17.9271	530.3830	600.6559
130	545.1020	17.9271	509.9656	580.2385
131	517.4954	17.9271	482.3589	552.6318
132	480.2759	17.9271	445.1395	515.4123
133	466.9840	17.9271	431.8476	502.1205
134	481.2010	17.9271	446.0646	516.3374
135	481.7638	17.9271	446.6274	516.9002
136	488.1042	17.9271	452.9677	523.2406
137	522.0907	17.9271	486.9543	557.2272
138	545.2331	17.9271	510.0967	580.3696
139	555.3350	17.9271	520.1986	590.4715
140	569.2723	17.9271	534.1358	604.4087
141	565.5195	17.9271	530.3830	600.6559
142	545.1020	17.9271	509.9656	580.2385
143	517.4954	17.9271	482.3589	552.6318
144	480.2759	17.9271	445.1395	515.4123



Here we can see the confidence intervals remain constant and the standard errors remain constant through out the forecast. We see a repeating periodic pattern around every 12 months which we also see in the original data.

SAS code:

```
filename repair "/home/u63378438/my_shared_file_links/huffer/repair.txt";

run;


data repair;

infile repair;

input x;

y=log(x);

run;




proc ARIMA data=repair plots=all;

identify var=x(1,12) nlag=24;

  estimate p=1 method=ml;

run;


proc ARIMA data=repair plots=all;

identify var=x(1,12) nlag=24;

estimate p=1 method=ml;

run;


proc ARIMA data=repair plots=all;

identify var=y(1,12) nlag=24;

estimate p=(1,12) q=(1) method=ml noconstant;

forecast lead=36 out=resids;
```

```
run;

proc ARIMA data=repair plots=all;
identify var=y(1,12) nlag=24;
estimate p=(1,12) q=(1) method=ml;
forecast lead=36 out=resids;
run;

filename bus "/home/u63378438/my_shared_file_links/huffer/bus.txt";
run;

filename bus "/home/u63378438/my_shared_file_links/huffer/bus.txt";
run;

data bus;
infile bus;
input x;
y=log(x);
run;

proc arima data=bus plots=all;
identify var=y(1, 12) nlag=24;
estimate q=(1,12) method=ml;
run;

proc arima data=bus plots=all;
identify var=x(1, 12) nlag=24;
estimate method=ml;
run;
```

```
proc arima data=bus plots=all;
identify var=y(1, 12) nlag=24;
estimate q=(1,12) method=ml;
forecast lead=36 out=resids;
run;

data oscale;
keep osforecast osL95 osU95;
set resids;
osforecast=exp(forecast);
osL95=exp(L95);
osU95=exp(U95);
run;



proc print data=oscale(firstobs=101);
run;

filename bus "/home/u63378438/my_shared_file_links/huffer/bus.txt";
run;

data bus;
infile bus;
input x;
y=log(x);
run;
```

```
proc arima data=bus plots=all;
identify var=y(1, 12) nlag=24;
estimate q=(1,12) method=ml;
run;

proc arima data=bus plots=all;
identify var=x(1, 12) nlag=24;
estimate method=ml;
run;

proc arima data=bus plots=all;
identify var=y(1, 12) nlag=24;
estimate q=(1,12) method=ml;
forecast lead=36 out=resids;
run;

data oscale;
keep osforecast osL95 osU95;
set resids;
osforecast=exp(forecast);
osL95=exp(L95);
osU95=exp(U95);
run;

proc print data=oscale(firstobs=101);
run;
```