Time Series Project

Reference Number: 22247197

Name: Siu Leung Lee

Email: vincent.lsl@gmail.com

Monthly Water Usage in Ontario, London

Introduction

The purpose of this time series project is to analyze the monthly water usage in Ontario, London from 1976 to 1987 and fit the series with ARIMA model. It is expected that the monthly water usage is highly correlated with the water usage in the preceding months and seasonal pattern is expected to be found in the data.

The project is divided into several parts. To begin with, the data will be examined with time series plot and correlogram in order to identify if there is any trend and seasonal pattern within the series. Second, attempts will be made on removing identified trend and seasonality and tentative model will be assumed based on sample autocorrelation of the transformed stationary time series. Third, different model dialogistic measures on residual like normal probability plot and histogram of residual will be performed for the tentative model so as to decide if the tentative one is the most suitable model to be used. Last but least, a forecast based on the most appropriate model will be check against actual data with a view to testing the validity of the model.

Data

The data used in this analysis is the actual monthly water usage (Ml/Day) in London, Ontario of the period from December 1976 to December 1987 with a total of 131 data points.

For the purpose of assure the validity of the proposed model, monthly data from Year 1976 to 1986 will be used to fit into a model but Year 1987 will be reserved for the evaluation of our model in ex-post forecast.

The data is coming from Time Series Data Library and can be found with the following link: http://robjhyndman.com/tsdldata/monthly/wqlondon.dat

The exhibits attached below show the time series plot and correlogram for the original time series. According to exhibit 1, it is apparent that the monthly water usage in Ontario, London is not stationary. First, it demonstrates an upwards linear trend in monthly water usage over the years. Second, there is a recurring pattern or seasonality within each year (i.e., a clear seasonal pattern with high water usage in July and low water usage in December).

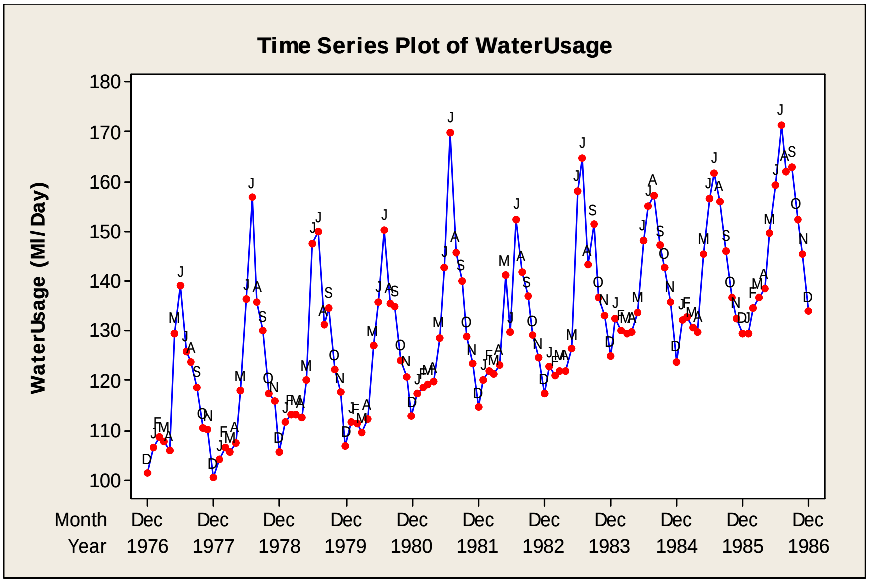

Exhibit 1 Time series plot of monthly water usage

Exhibit 2 further confirms the seasonal relationship within the time series where seasonal autocorrelation relationships can be found notably. There are strong correlations at lags 12, 24 and so on.

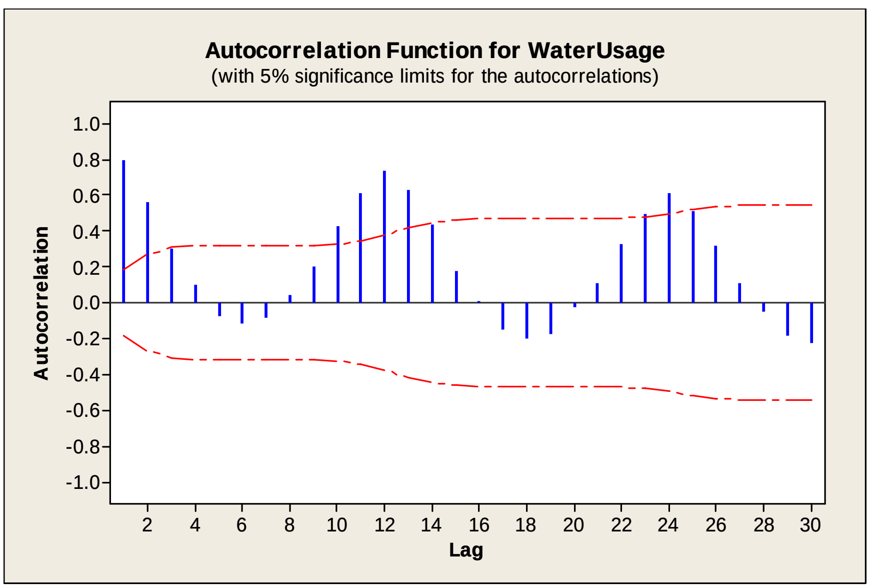

Exhibit 2 Sample ACF of time series of monthly water usage

Because of the fact that both the upward trend and seasonality mentioned above will disrupt our analysis and hence deseasonalizing should be performed before any further analyses.

With a view to removing the upward trend on the time series, 1st difference (Y_t-Y_{t-1}) of the time series is performed and the time series plot is shown as follow:

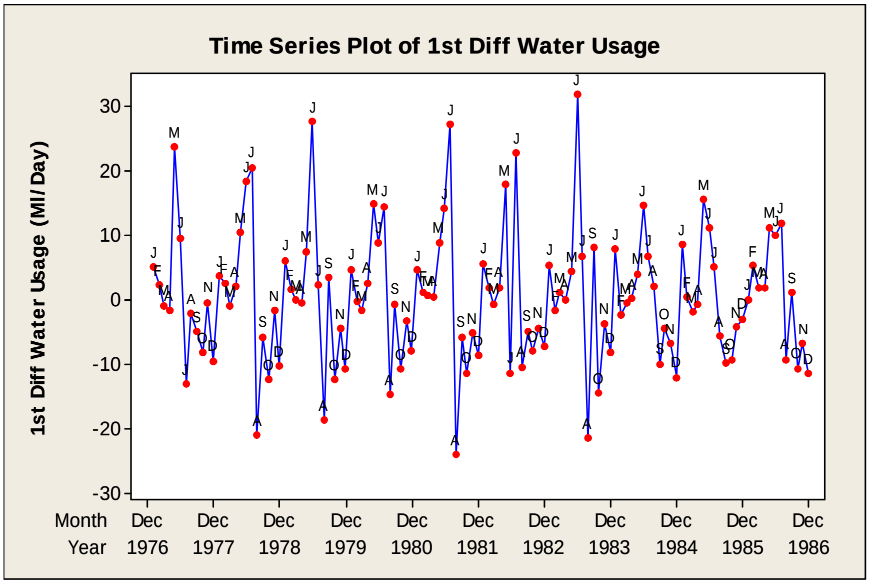

Exhibit 3 Time series plot of 1st differences of monthly water usage

According to exhibit 3, the upward trend of the time series is taken away and looks like having a constant mean. However, the time series has to be deseasonlized before fitting into time series model. Given a strong annual pattern can be found from exhibit 2, the time series will be transformed into difference of water usage in the current month and the same month in the previous year (Y_t-Y_{t-12}). The resulting time series plot is shown as follow:

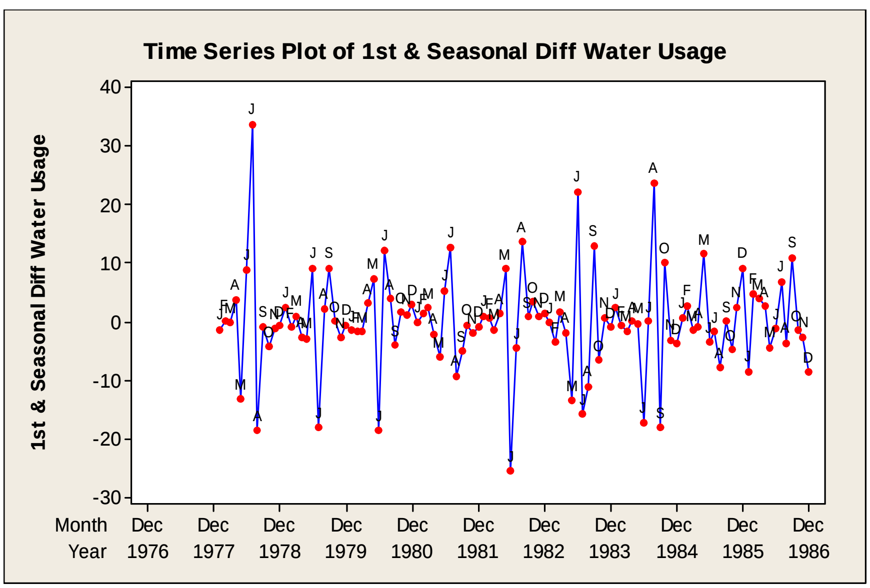

Exhibit 4 Time series plot of 1st and seasonal differences of monthly water usage

Exhibit 4 displays the time series plot of the monthly water usage after taking both a 1st and seasonal differences. It appears that most, if not all, of the seasonality is gone now.

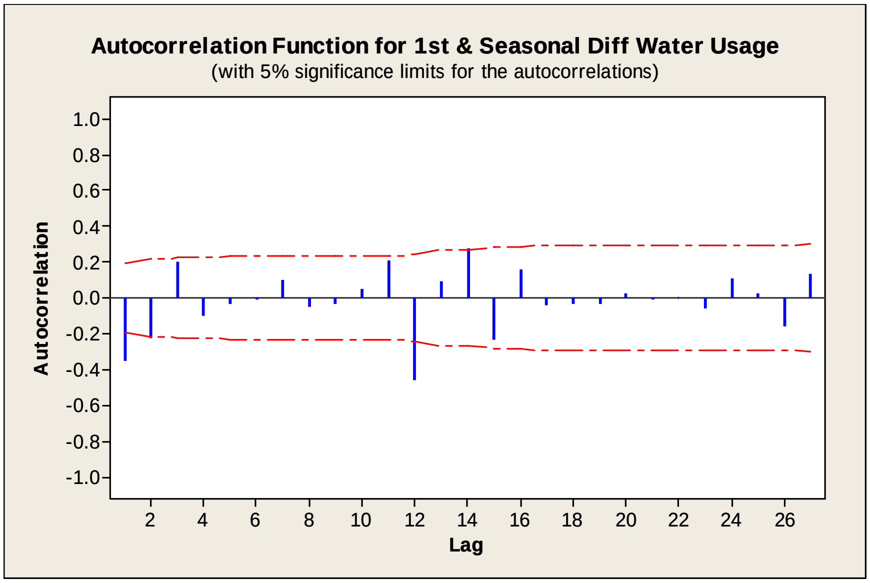

Exhibit 5 Sample ACF of 1st and seasonal differences of monthly water usage

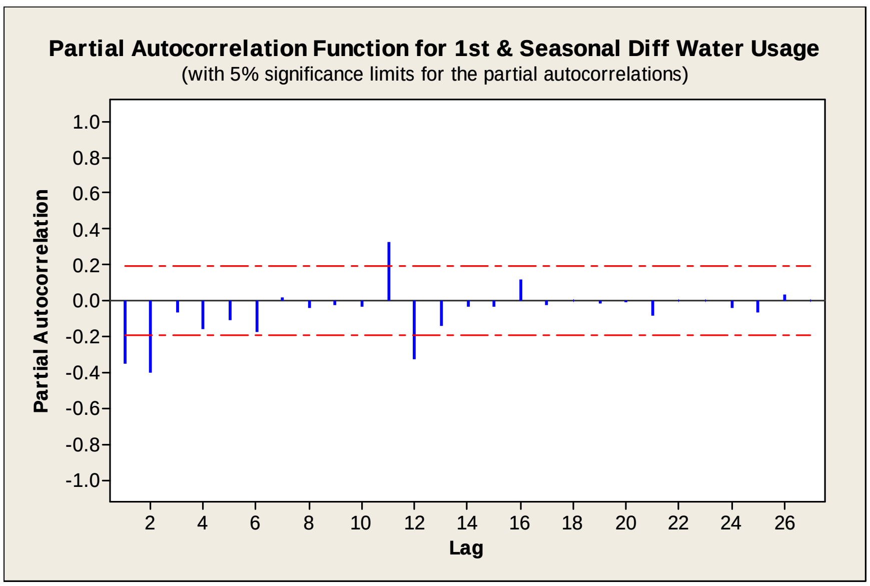

Exhibit 6 Sample PACF of 1st and seasonal differences of monthly water usage

Exhibit 5 and 6 further verifies that very little autocorrelation remains in the time series after both 1st and seasonal differences being performed. This plot also suggested that a time series model with lag 1 and lag 12 autocorrelations may be adequate.

Model Fitting

As mentioned in the previous session exhibit 5, the sample autocorrelation function suggests the possibility of using a time series model that is able to incorporate both lag 1 and lag 12. After fitting with model with different AR and MA parameters, one of the models that can incorporate the above requirements is the multiplicative, seasonal ARIMA model (2, 1, 1) X $(1, 1, 1)_{12}$ model and it will also be used as tentative model that will undergo further analyses in the latter session.

ARIMA (2, 1, 1) X $(1, 1, 1)_{12}$:

$$W_t = \mu + \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \Phi W_{t-12} - \Phi\varphi_1 W_{t-13} - \Phi\varphi_2 W_{t-14} + e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}$$

$$\text{where } W_t = \nabla_{12}\nabla Y_t$$

Exhibit 7 shows the estimation of parameter of the ARIMA (2, 1, 1) X $(1, 1, 1)_{12}$ model and their standard errors. It can be found that $\varphi_1 = -0.5764$, $\varphi_2 = -0.3613, \Phi = -0.6684, \theta = -0.9075, \Theta = 0.7853, \mu = 0.01789$ and the model become.

ARIMA (2, 1, 1) X $(1, 1, 1)_{12}$:

$$W_t = 0.01789 - 0.5764 W_{t-1} - 0.3613 W_{t-2} - 0.6684 W_{t-12} - 0.3853 W_{t-13} - 0.2415 W_{t-14} + e_t + 0.6684 e_{t-1} - 0.7853 e_{t-12} - 0.7127 e_{t-13}$$

$$\text{where } W_t = \nabla_{12}\nabla Y_t$$

Since the parameter is estimated with high significant, the proposed model will be proceeded to diagnostic check.

ARIMA Model: 1st & Seasonal Diff Water Usage

Estimates at each iteration

Iteration	SSE	Parameters					
0	55331.6	0.100	0.100	0.100	0.100	0.100	-0.091
1	50368.3	-0.050	0.052	0.089	0.005	0.111	-0.108
2	47986.8	-0.200	0.020	0.195	-0.118	0.231	-0.109
3	46556.2	-0.350	-0.008	0.241	-0.252	0.287	-0.116
4	45481.6	-0.500	-0.033	0.257	-0.391	0.311	-0.127
5	44644.5	-0.650	-0.058	0.265	-0.534	0.326	-0.139
6	43995.7	-0.800	-0.082	0.269	-0.679	0.336	-0.151
7	43651.3	-0.950	-0.103	0.269	-0.826	0.340	-0.164
8	42419.6	-0.807	-0.095	0.268	-0.676	0.353	-0.151
9	39927.9	-0.675	-0.100	0.265	-0.526	0.379	-0.138
10	35224.9	-0.567	-0.131	0.220	-0.376	0.389	-0.134
11	31592.3	-0.462	-0.155	0.213	-0.226	0.430	-0.121
12	28049.2	-0.367	-0.179	0.196	-0.076	0.462	-0.107
13	22039.7	-0.470	-0.270	0.220	-0.082	0.612	-0.092
14	18523.1	-0.408	-0.301	0.152	0.068	0.625	-0.081
15	15525.2	-0.356	-0.324	0.089	0.218	0.647	-0.065
16	12957.9	-0.315	-0.339	0.028	0.368	0.674	-0.046
17	9596.4	-0.314	-0.366	-0.062	0.518	0.787	-0.048
18	8001.1	-0.297	-0.360	-0.134	0.668	0.807	-0.017
19	6638.6	-0.299	-0.351	-0.210	0.818	0.827	0.008
20	5838.5	-0.388	-0.368	-0.360	0.937	0.813	0.022
21	5147.5	-0.522	-0.408	-0.510	0.920	0.797	0.034
22	4983.8	-0.571	-0.393	-0.589	0.901	0.783	0.013

23	4949.7	-0.571	-0.371	-0.630	0.912	0.782	0.019
24	4936.7	-0.581	-0.373	-0.654	0.899	0.785	0.014
25	4931.4	-0.576	-0.361	-0.668	0.907	0.785	0.018

** Convergence criterion not met after 25 iterations **

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	-0.5764	0.1062	-5.43	0.000
AR	2	-0.3613	0.1058	-3.41	0.001
SAR	12	-0.6684	0.0859	-7.78	0.000
MA	1	0.9075	0.0640	14.18	0.000
SMA	12	0.7853	0.0960	8.18	0.000
Constant		0.01789	0.02275	0.79	0.434

Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 108, after differencing 95
Residuals: SS = 3970.59 (backforecasts excluded)
MS = 44.61 DF = 89

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	10.3	28.3	44.8	56.0
DF	6	18	30	42
P-Value	0.112	0.057	0.041	0.073

Forecasts from period 121

Period	Forecast	95% Limits Lower	95% Limits Upper	Actual
122	7.6511	-5.4430	20.7453	
123	-1.6318	-16.1784	12.9148	
124	-2.7148	-17.2620	11.8325	
125	-0.4976	-15.4425	14.4474	
126	3.9039	-11.0630	18.8708	
127	-4.4890	-19.4623	10.4844	
128	-3.3974	-18.4228	11.6281	
129	3.7772	-11.2520	18.8063	
130	-6.1820	-21.2219	8.8578	
131	1.4907	-13.5686	16.5499	
132	1.4344	-13.6354	16.5043	
133	6.7106	-8.3715	21.7928	

Exhibit 7 Parameter Estimates for monthly water usage with ARIMA (2, 1, 1) X $(1, 1, 1)_{12}$ model

Model Diagnostic

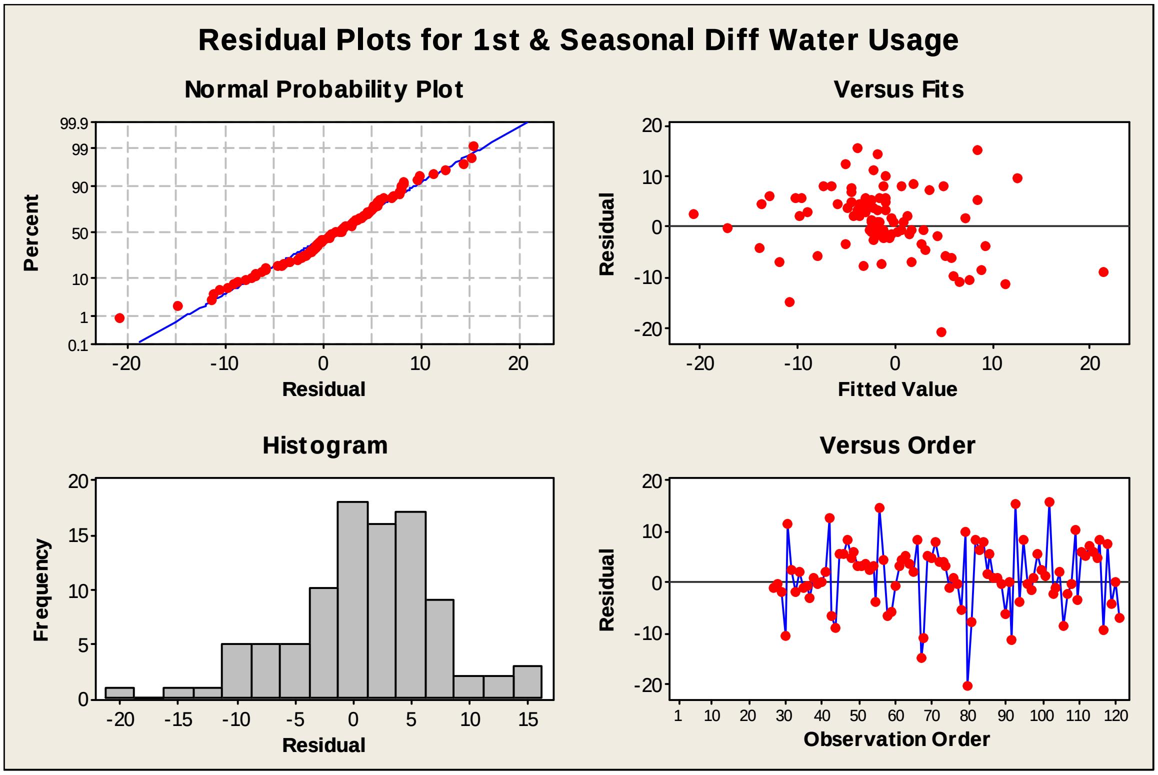

Exhibit 8 Residual plots for monthly water usage with ARIMA (2, 1, 1) X (1, 1, 1)12 model

When we look at the time series plot of the residuals (bottom right corner of the residual plots), the plot does not suggest any main irregularities within the model and the residuals are nearly uncorrelated with each other. It is worth mentioning that there is a large residual in July of 1983 which suggests that the time series may contain outliner.

Besides, the bell-shaped histogram and the normal probability plot on the residual also support that ARIMA (2, 1, 1) X (1, 1, 1)12 model is a appropriate model of this time series.

Furthermore, exhibit 9 and 10 ACF and PACF of residual for 1st and seasonal difference water usage also shows that there is no significant correlation within the sample time series with 5% significance limits.

Last but not least, the Modified Box-Pierce (Ljung-Box) Chi-Square statistic shown in exhibit 7 illustrates that ARIMA (2, 1, 1) X (1, 1, 1)12 model passed the test with 5% significant level which means that the null hypothesis that the residuals follow a white noise process is not rejected at 5% significant level.

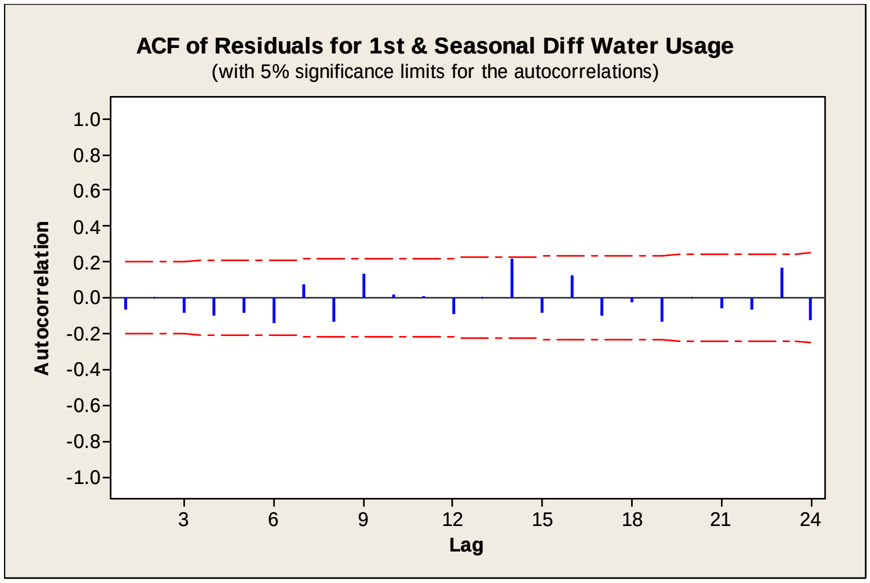

Exhibit 9 ACF of residual for 1st and seasonal difference water usage

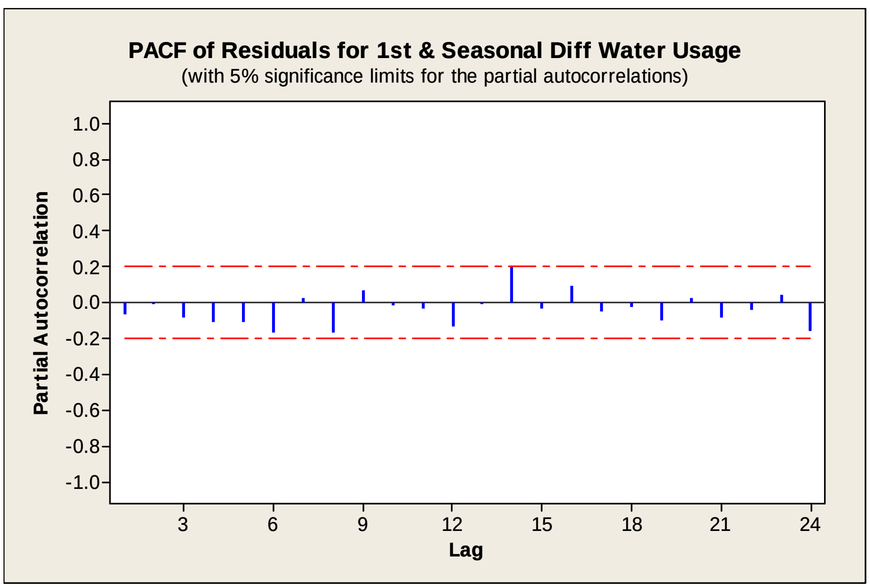

Exhibit 10 PACF of residual for 1st and seasonal difference water usage

Model Evaluation

After identifying and verifying the ARIMA (2, 1, 1) x $(1, 1, 1)_{12}$ model, the forecasted monthly water usage will be tested against the existing data in 1987. After restoring back the deseaonalized 1st difference stationary model into original time series, the time series plot of forecasted data and original data is shown below.

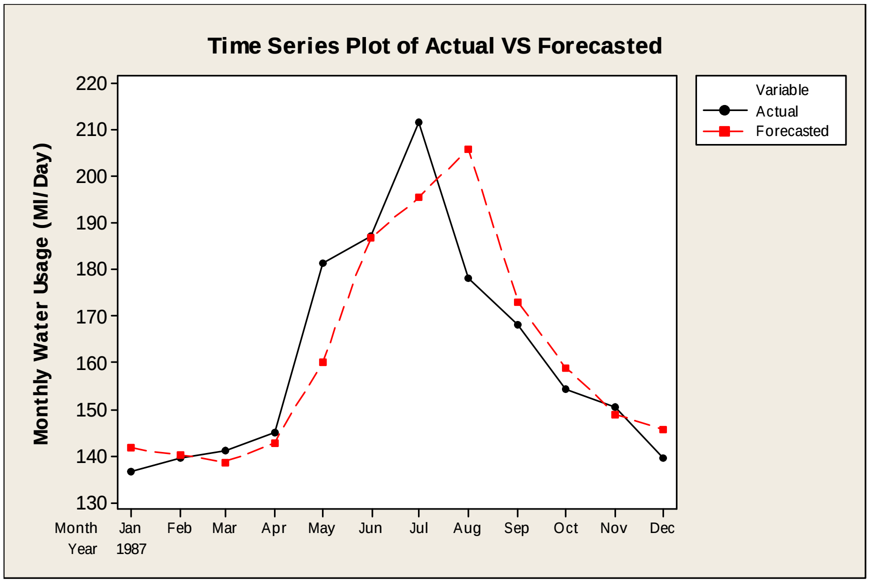

Exhibit 11 Time series plot of actual vs forecasted time series during 1987

As we can see form the plot above, although there are some spikes, ARIMA (2, 1, 1) x $(1, 1, 1)_{12}$ does pretty good in predicting the monthly water usage.

Conclusion

We have performed examination on the times series of monthly water usage in Ontario, London. After 1st differences and deseasonalization, the time series is transformed into a stationary time series. The transformed series is fitted with multiplicative, seasonal ARIMA (2, 1, 1) x $(1, 1, 1)_{12}$ model with equation:

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

$$W_t = 0.01789 - 0.5764W_{t-1} - 0.3613W_{t-2} - 0.6684W_{t-12} - 0.3853W_{t-13} - 0.2415W_{t-14} + e_t + 0.6684e_{t-1} - 0.7853\mathrm{e}_{t-12} - 0.7127\mathrm{e}_{t-13}$$

$$\text{where } W_t = \nabla_{12}\nabla Y_t$$

Residuals analyses have also be performed in order to check the validity of the model. No main irregularities within the model are found and the residuals are nearly uncorrelated with each other. Furthermore, forecasted monthly water usage is tested against the actual so as to evaluate the effectiveness of the model and satisfactory results are also being shown on the test. Overall, it can be concluded that the monthly water usage in Ontario, London can be fit with multiplicative, seasonal ARIMA (2, 1, 1) x $(1, 1, 1)_{12}$ model.