5/13/2020

Forecast Daily Interstate 94 Westbound Traffic Volume for MN DoT ATR Station 301

Jennifer Nguyen, Guotian Kan,
Viktoriia Dunleavy, Yuzhou Ding,
Ci Song

Table of Content

1. Introduction and Overview

Data set details and variable definitions:

The data is about Daily Interstate 94 Westbound traffic volume for MN DoT ATR station 301, roughly midway between Minneapolis and St Paul, MN. The daily weather features and holidays are included to assess impacts on traffic volume. The data set has 1004 observations from 1/1/2016 to 9/30/2018. The data set comes from Social Data Science & General Mills.

The data set has 7 variables. Below is the variable information:

Main variable of interest/response variable: **sum_traffic_volume** (Numeric): Daily I-94 ATR 301 reported westbound traffic volume

Other variables:

date_time (DateTime): Day of the data collected in local CST time

Variables to use in univariate time series model:

holiday(Categorical): US National holidays plus regional holiday, Minnesota State Fair

Day_of_week: the name of the date (Mon, Tue, Wed, Thu, Fri, Sat, Sun)

Variables to use in multivariate time series model:

average_temp (Numeric): Average temp in kelvin

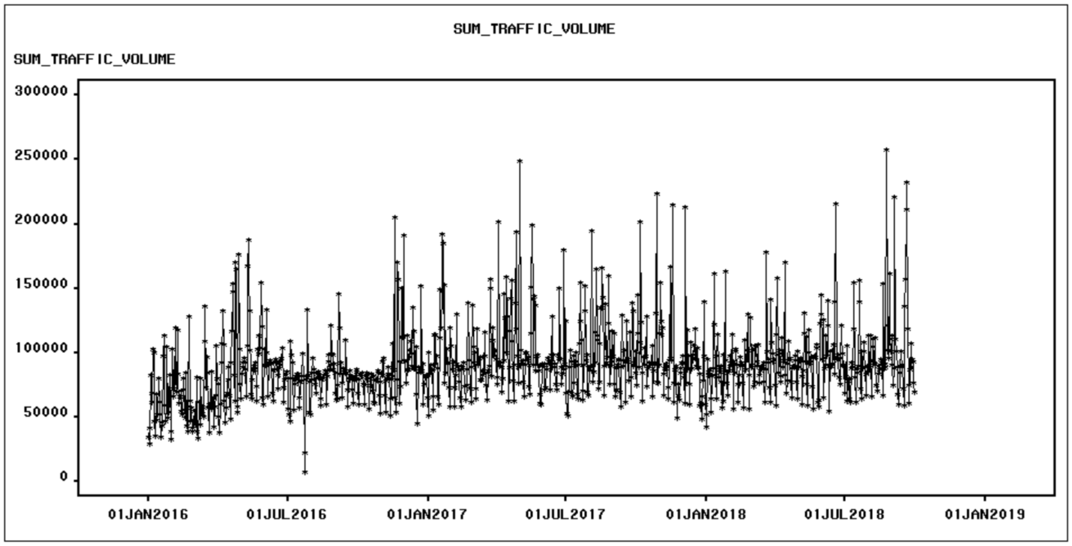
sum_rain_1h (Numeric):Amount in mm of rain that occurred that day

average_clouds_all (Numeric): Average percentage of cloud cover

The interest of data set and goal of this project:

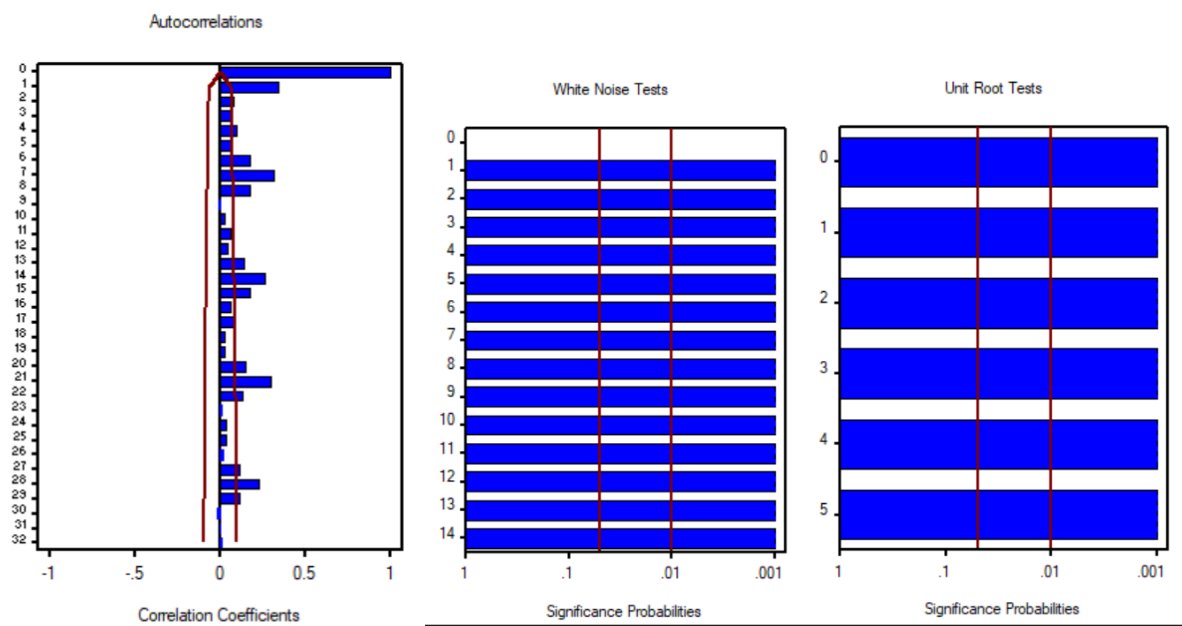
The data set could be used to analyze the relationship between traffic volume and its relationship with time, represented by univariate deterministic and stochastic models, as well as to analyze the effect of multiple predictors on traffic volume, represented by multivariate time series models. It is interesting to use Time series analysis to reduce the residuals to white noise and capture more information from data. Whether it has seasonal patterns, linear trends or other relationships such as AR and MA. In this project, we will use different methods to analyze the data to find out the most suitable model to fit the data and predict future observations of traffic volume.

Now, let's look at the details of the main series sum_traffic_volume.



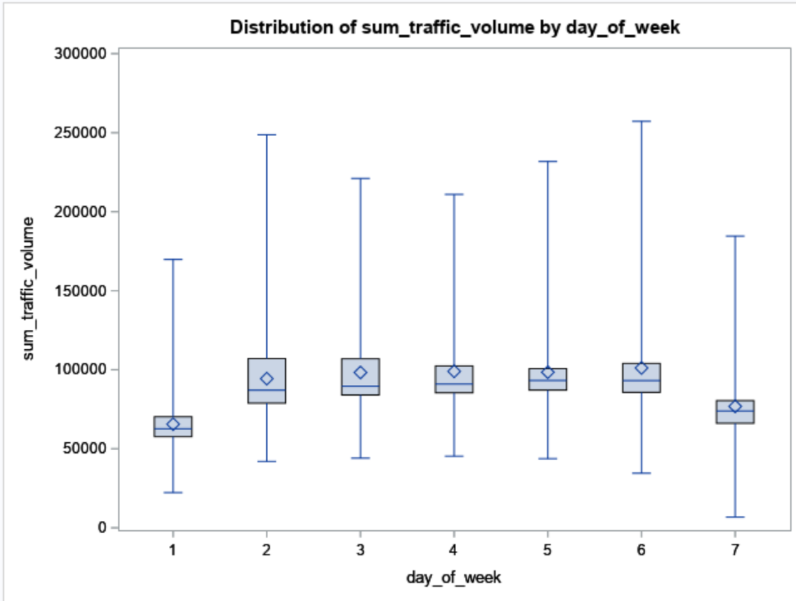
Output 1.1.

According to the Output 1.1 the time series do not have upward or downward trends. However, the time series shows the signs of seasonality. The interval between peaks and troughs are similar in length as we can observe in this daily data set.



Output 1.2.

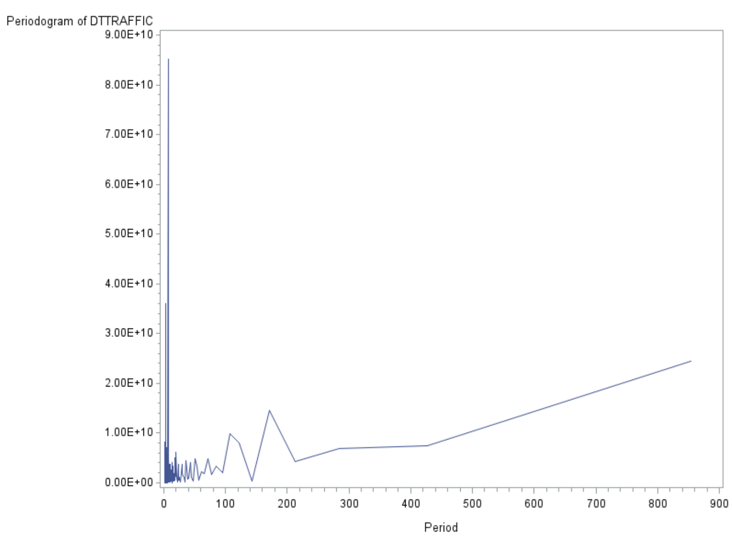
Looking at the Output 1.2, the Autocorrelation plot shows that the time series is not a white noise. The White Noise Test confirmed that the time series is not a white noise series. Based on the Unit Root Tests, we can conclude that the time series is non-seasonal stationary. However, the sample autocorrelation function (ACF) shows signs of seasonality as the series decays slowly over seasonal lags such as lag 7, lag 14, and lag 21. Thus, the series looks non-stationary at seasonal lags.



Output 1.3

(1 = Sun, 2 = Mon, 3 = Tue, 4 = Wed, 5 = Thu, 6 = Fri, 7 = Sat)

Output 1.3 shows the boxplots of the time series. The observation is that Sundays followed by Saturdays are likely to be the least busy days of the week for traffic in our dataset. The medians of the traffic volume across weekdays from Monday to Friday look comparable with no significant difference.



Output 1.4.

Since the time series show the signs of seasonality we used Periodogram to capture it. According to the Output 1.4 we identify that there are some hidden periodicities of seasonality. In the Output 1.5 we combined 6 harmonics with the highest amplitudes. These harmonics are 122, 244, 1, 5, 8, and 366, and their respective corresponding periods are 7, 3.5, 854, 170.8, 106.75, and 2.333. To keep the model parsimonious, we would like to only use 6 harmonics in the cyclical trend model later.

Obs	FREQ	PERIOD	P_01
123	0.8976	7	85108
245	1.7952	3.5	36098
2	0.00736	854	24422
6	0.03679	170.8	14586
9	0.05886	106.75	9843
367	2.69279	2.333	8209

Output 1.5.

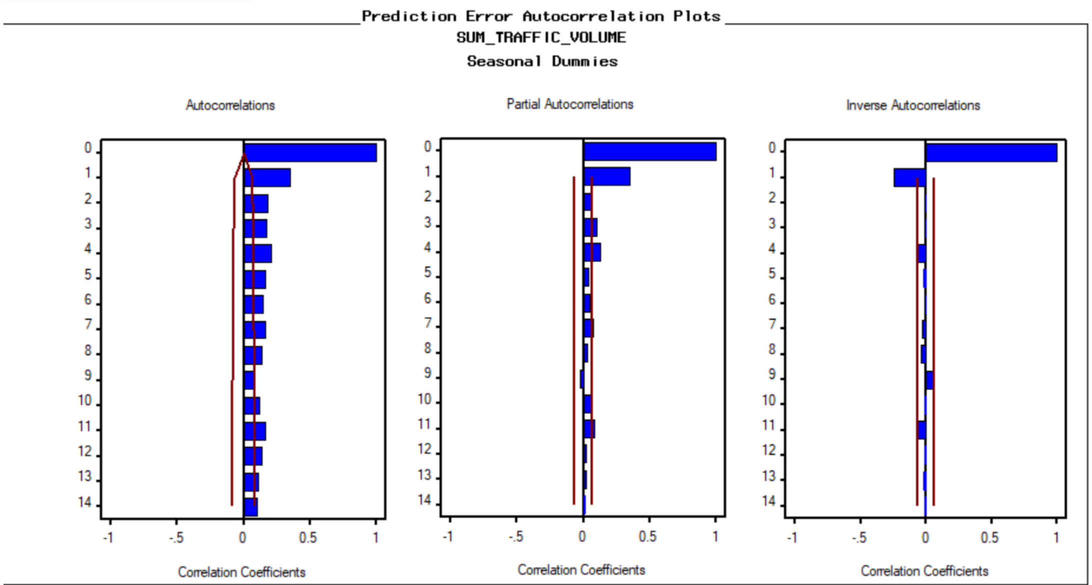
In the following sections, we will be fitting our time series data to different models with a hold-out period including 150 observations (about 15% of the total number of observations).

2. Univariate Time Series Models

2.1. Deterministic Time-series models and Error model

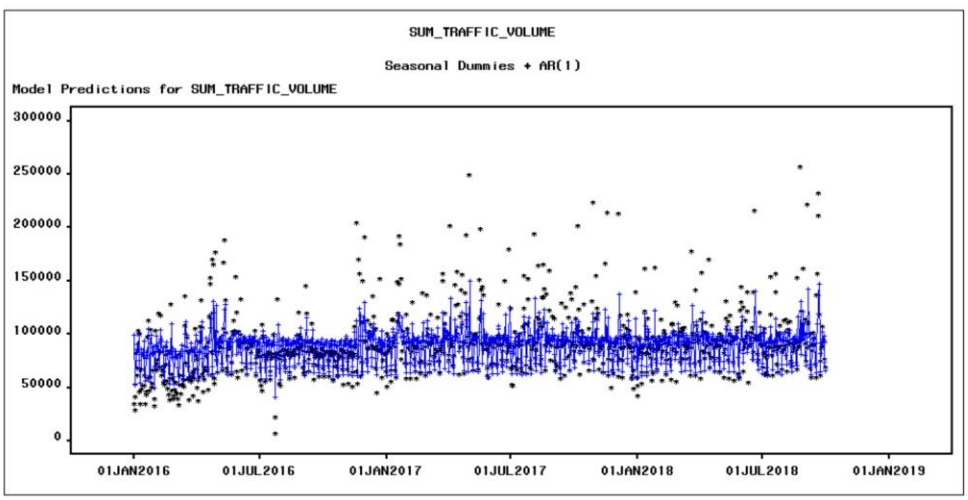
The response variable in this dataset is a daily traffic volume. Based on the graph of the real data, the sample ACF as well as commonly known we can safely assume that our data is seasonality. Therefore, we can build seasonal dummies models and a cyclical trend model, then determine the best seasonal model according to their test Root Mean Squared Error (RMSE). We are not fitting a trend model because our data has no signs of trend or non-seasonal nonstationarity.

The first model is a seasonal dummies only model. According to the Output 2.1 it decays gradually, while the PACF is chopped off to 0 after lag 1. This behaviour of the residuals indicates an AR(1) process . To further improve the model, we will apply the AR(1) process to the error term (an error model) in the seasonal dummies model.



Output 2.1

The model predictions of the Seasonal Dummies + AR(1) can be found in Output 2.2.



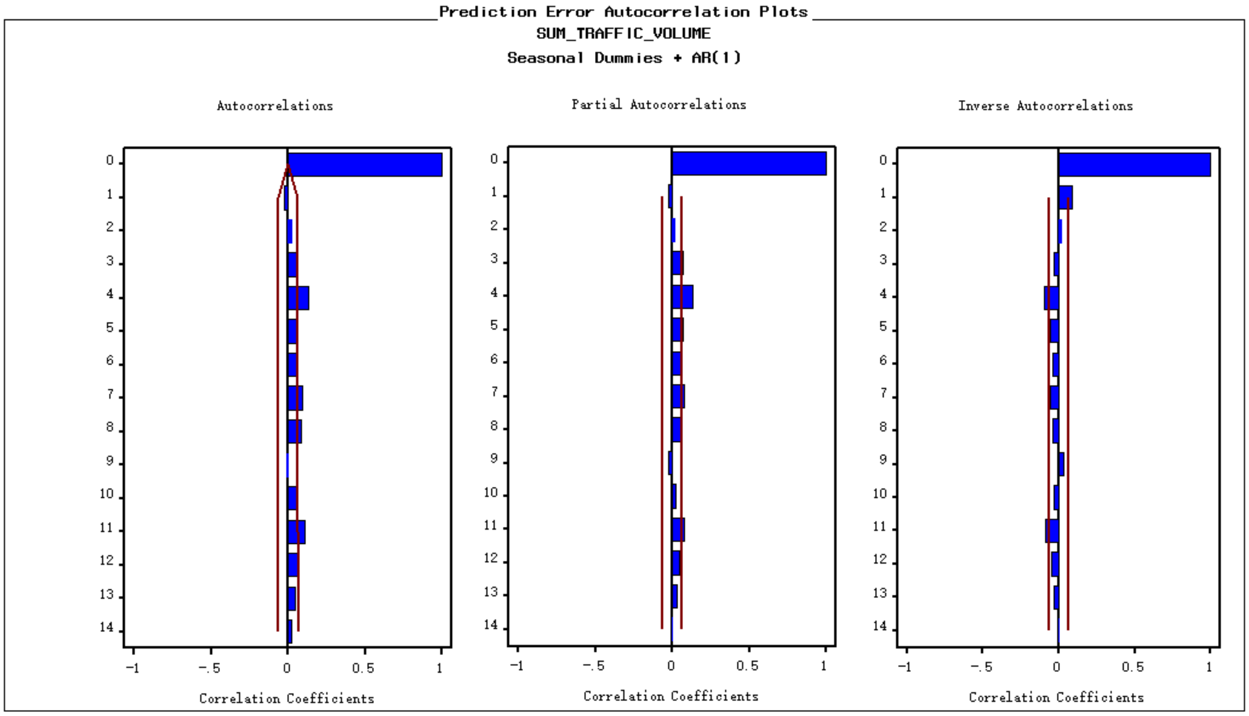
Output 2.2

Parameter Estimates
SUM_TRAFFIC_VOLUME
Seasonal Dummies + AR(1)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	75457	2401	31.4258	<.0001
Autoregressive, Lag 1	0.35242	0.0322	10.9558	<.0001
Seasonal Dummy 1	-10320	2729	-3.7813	0.0002
Seasonal Dummy 2	18204	3168	5.7455	<.0001
Seasonal Dummy 3	19833	3294	6.0209	<.0001
Seasonal Dummy 4	22573	3294	6.8519	<.0001
Seasonal Dummy 5	21269	3169	6.7111	<.0001
Seasonal Dummy 6	23786	2729	8.7166	<.0001
Model Variance (sigma squared)	615367462	.	.	.

Output 2.3

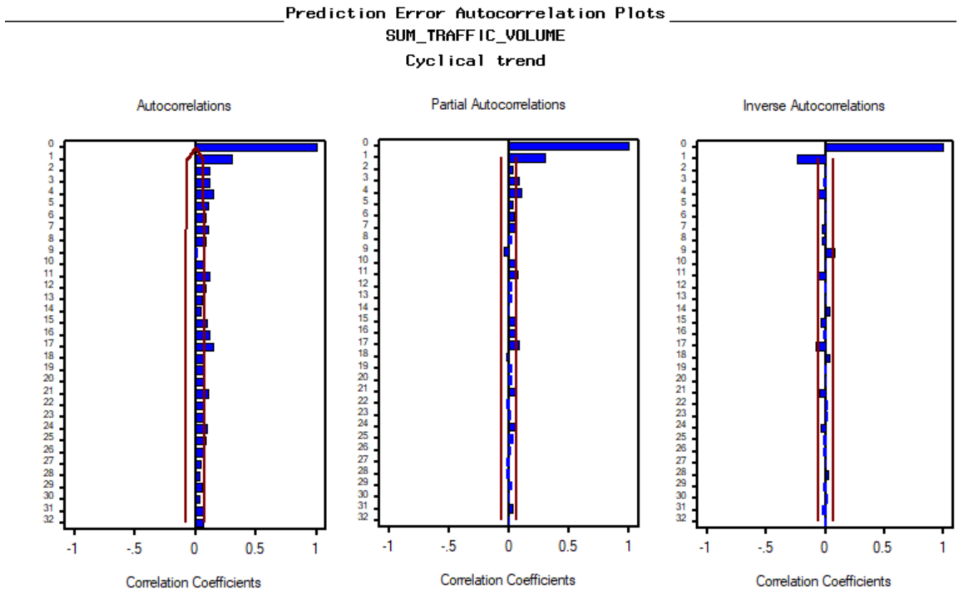
According to the Output 2.3, all independent variables are statistically significant at 5%. Referenced date is Saturday, Seasonal Dummy 1 is Sunday, Seasonal Dummy 2 is Monday, Seasonal Dummy 3 is Tuesday, Seasonal Dummy 4 is Wednesday, Seasonal Dummy 5 is Thursday, Seasonal Dummy 6 is Friday.



Output 2.4

The ACF of the seasonal dummies + AR(1) mode is presented in Output 2.4. It indicates that the errors of the model are white noise, despite the spikes at lag 4 and 11, which can be considered insignificant in this model. Therefore, this model captures all of the autocorrelations of the main series data.

In order to capture the hidden seasonality, we will build a cyclical trend model based on the six most prominent periodicities that we have identified in Output 1.5. For each of those periodicities, we created a sine-cosine pair based on the harmonics (122, 244, 1, 5, 8, and 366).



Output 2.5

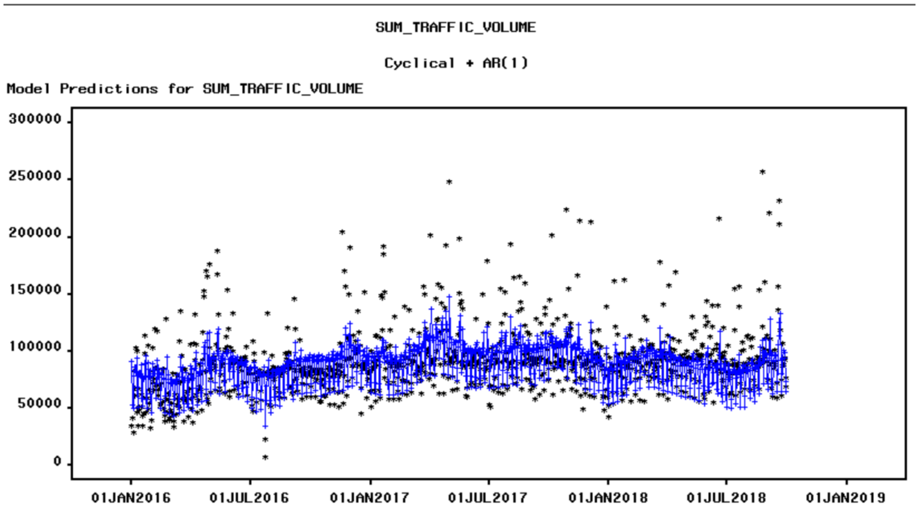
Based on Output 2.5, the ACF of the errors decays quickly over lags while PACF drops to 0 after lag 1. Thus, we will apply an AR(1) process tor the error terms to improve this model.

Parameter Estimates
SUM_TRAFFIC_VOLUME
Cyclical + AR(1)

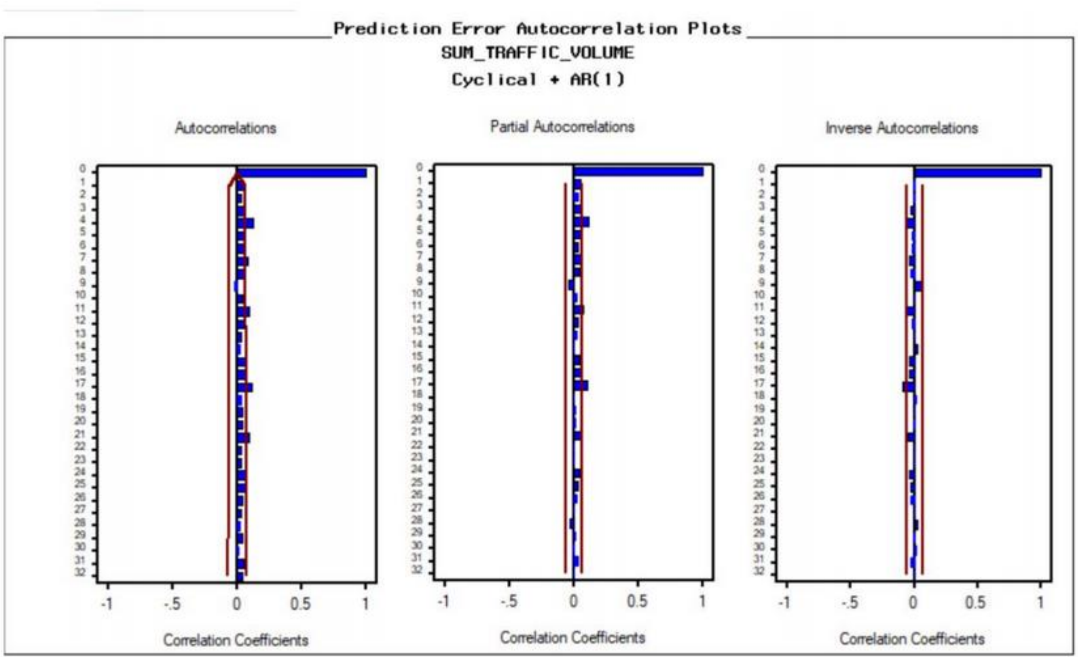
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	89101	1092	81.6051	<.0001
Autoregressive, Lag 1	0.25024	0.0334	7.4915	<.0001
COS122	11068	1336	8.2824	<.0001
SIN122	-8800	1337	-6.5824	<.0001
COS244	-552.43426	1069	-0.5167	0.6062
SIN244	9166	1069	8.5745	<.0001
COS1	-7562	1543	-4.8994	<.0001
SIN1	-7968	1545	-5.1580	<.0001
COS5	1774	1543	1.1496	0.2523
SIN5	-7070	1544	-4.5783	<.0001
COS8	-2459	1542	-1.5941	0.1132
SIN8	3192	1544	2.0678	0.0406
COS366	-2819	941.8096	-2.9937	0.0033
SIN366	-3362	941.4113	-3.5718	0.0005
Model Variance (sigma squared)	572796829	.	.	.

Output 2.6

According to Output 2.6, more than half of the independent variables are statistically significant. The predicted values are plotted against the actual values in Output 2.7



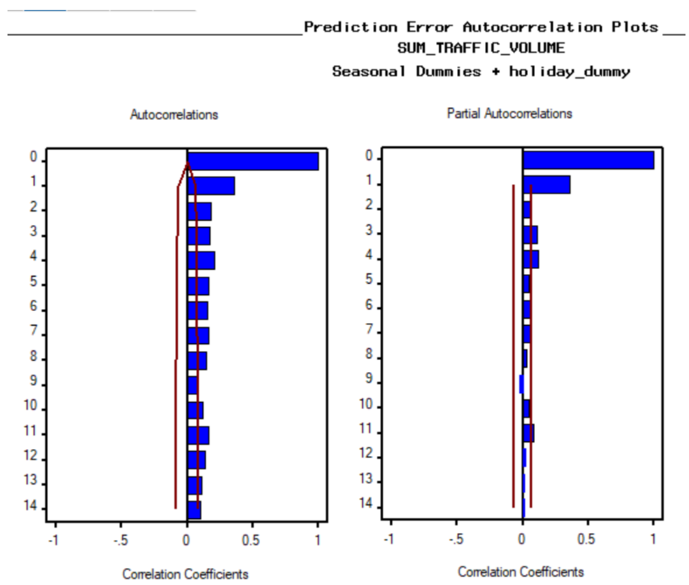
Output 2.7



Output 2.8

The autocorrelation plots of the errors of the cyclical trend + AR(1) model is presented in Output 2.8. The errors of the model are white noise, despite the spikes at lag 4, 11 and 17, which can be considered insignificant in this model.

In order to capture the impact of the holiday, we will add a Dummy variable, holiday_dummy, to the model. Based on the shape of the ACF (decays over lags) and PACF (chopped off after lag 1) in output 2.9d, we will use an AR(1) error model to make the residuals white noise.



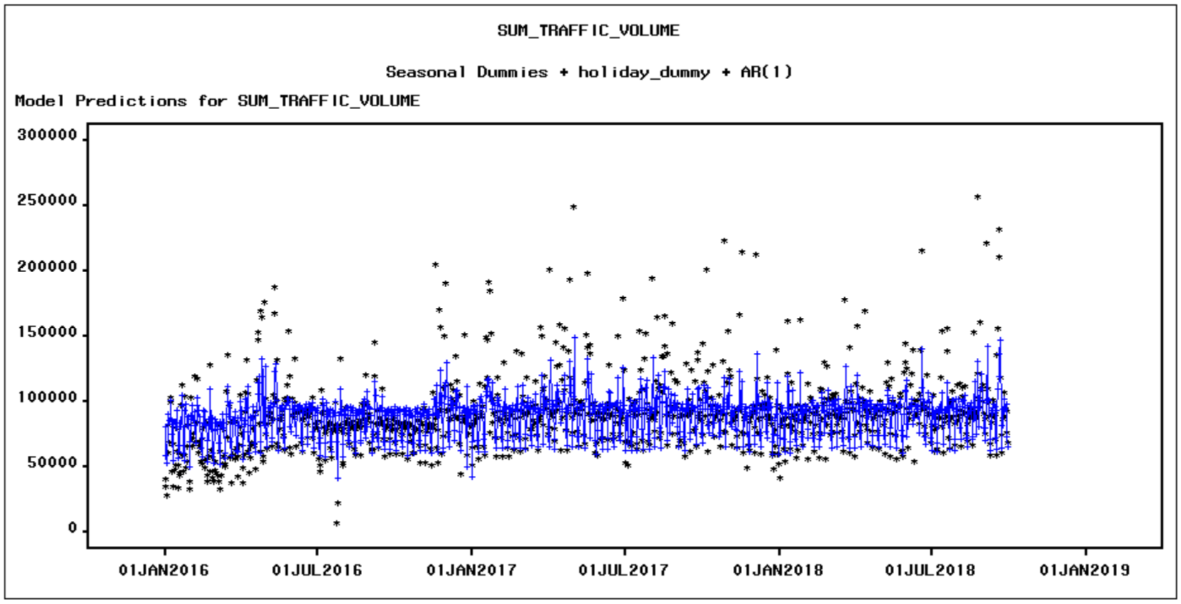
Output 2.9d

We will also use an AR(1) error model to make the residuals white noise. According to the Output 2.9a, the p-values of all the variables are less than 0.005, suggesting that the variables are statistically significant. The predicted values are plotted against the actual values in Output 2.9a. The dummy variable helped us to capture predicted values better than the previous models.

Parameter Estimates
SUM_TRAFFIC_VOLUME
Seasonal Dummies + holiday_dummy + AR(1)

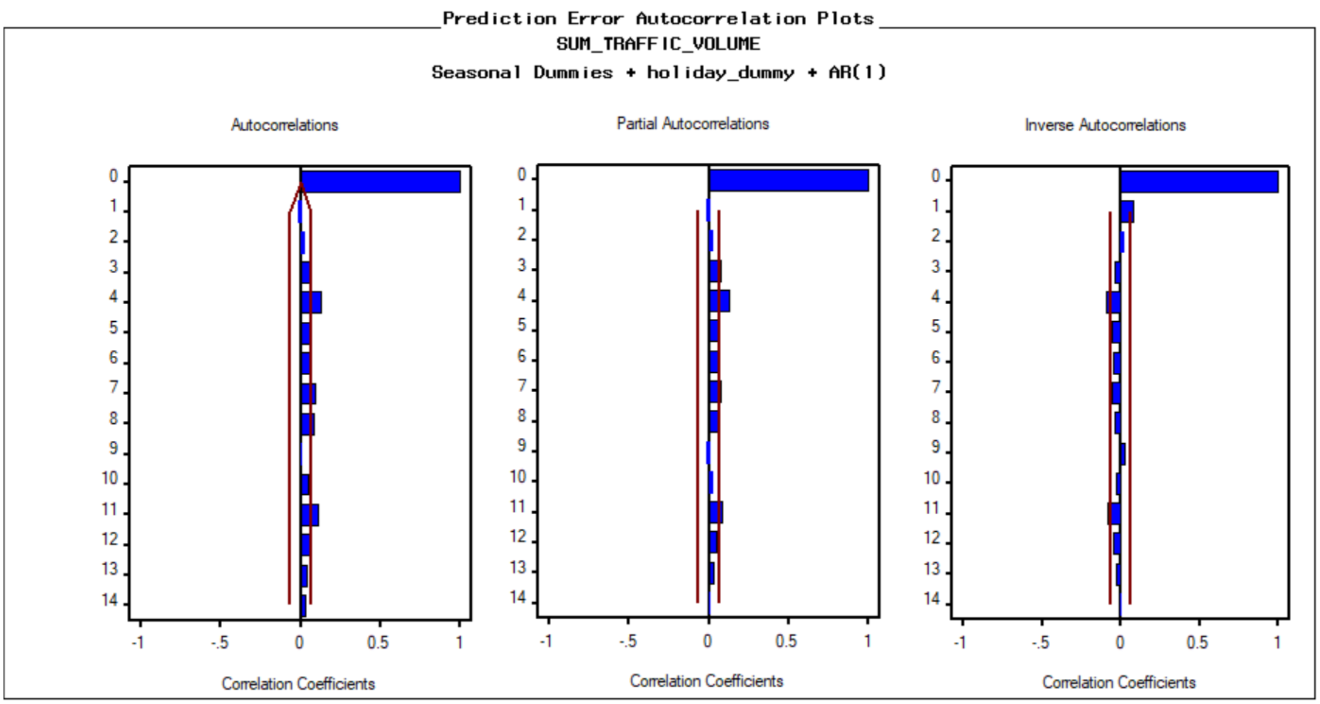
Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Intercept	75457	2380	31.7087	<.0001
Autoregressive, Lag 1	0.35079	0.0322	10.8898	<.0001
Seasonal Dummy 1	-10011	2709	-3.6946	0.0003
Seasonal Dummy 2	20520	3198	6.4162	<.0001
Seasonal Dummy 3	19994	3266	6.1212	<.0001
Seasonal Dummy 4	22592	3266	6.9164	<.0001
Seasonal Dummy 5	21940	3148	6.9698	<.0001
Seasonal Dummy 6	24247	2711	8.9456	<.0001
holiday_dummy	-18817	4833	-3.8934	0.0002
Model Variance (sigma squared)	605238917	.	.	.

Output 2.9a



Output 2.9b

According to the Output of 2.9c, The errors of the model are white noise, despite the spikes at lag 4 and 11, which can be considered insignificant in this model.



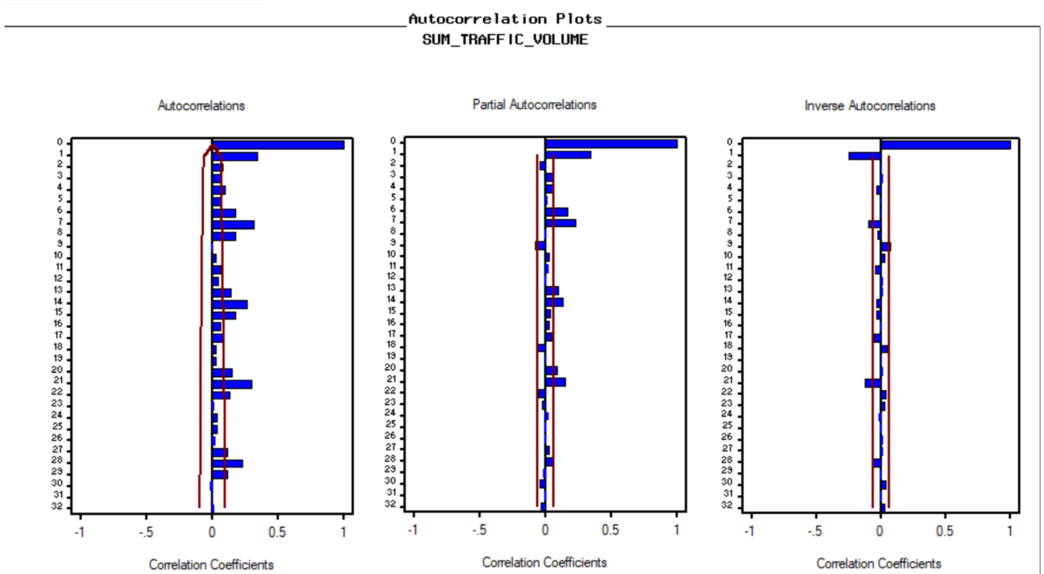
Output 2.9c

According to the results we get from each model shown in output 2.9e, we can get the RMSE of Seasonal Dummies + Holiday Dummy + AR(1) is 27928.4, which is the smallest among all the models in this part. We have tried several different models and the results indicate that the seasonal trend exists and whether the day is holiday does impact the volume of the traffic.

		Test Error	
	Model Fit (Square Root of Model Variance)	Validation RMSE	Validation MAPE
Seasonal Dummies	26482.6	30550.0	14.035
Seasonal Dummies + AR(1)	24806.6	28615.1	14.690
Seasonal Dummies + Holiday Dummy + AR(1)	24601.6	27928.4	14.462
Cyclical trend	24701.6	34651.9	18.092
Cyclical trend + AR(1)	23933.2	31291.8	15.404

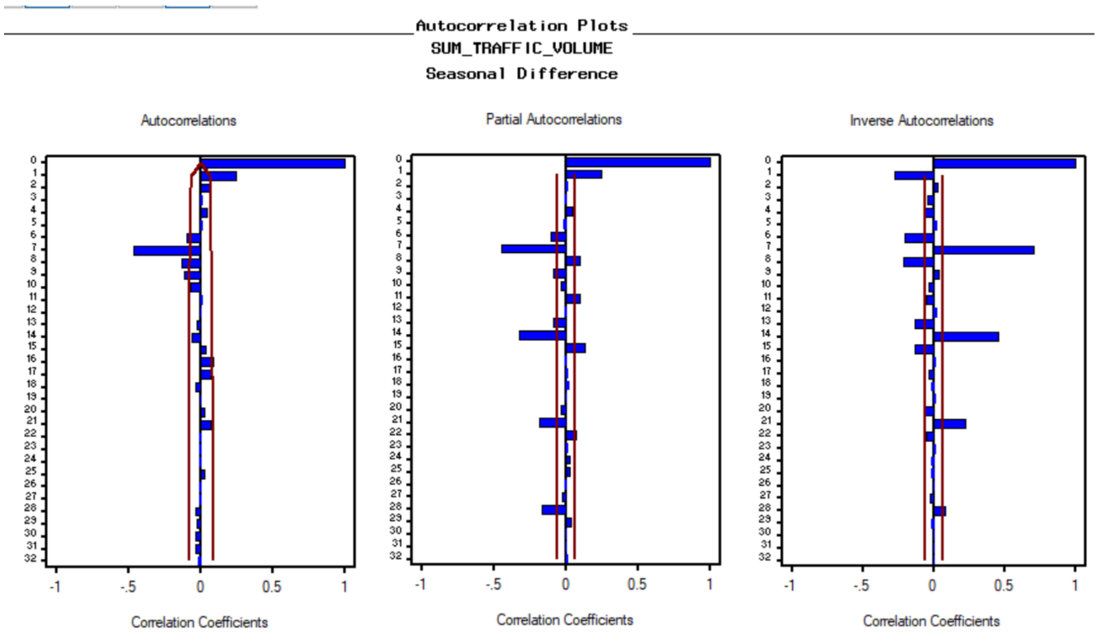
Output 2.9e

2.2. ARIMA Models with Seasonal ARIMA Components



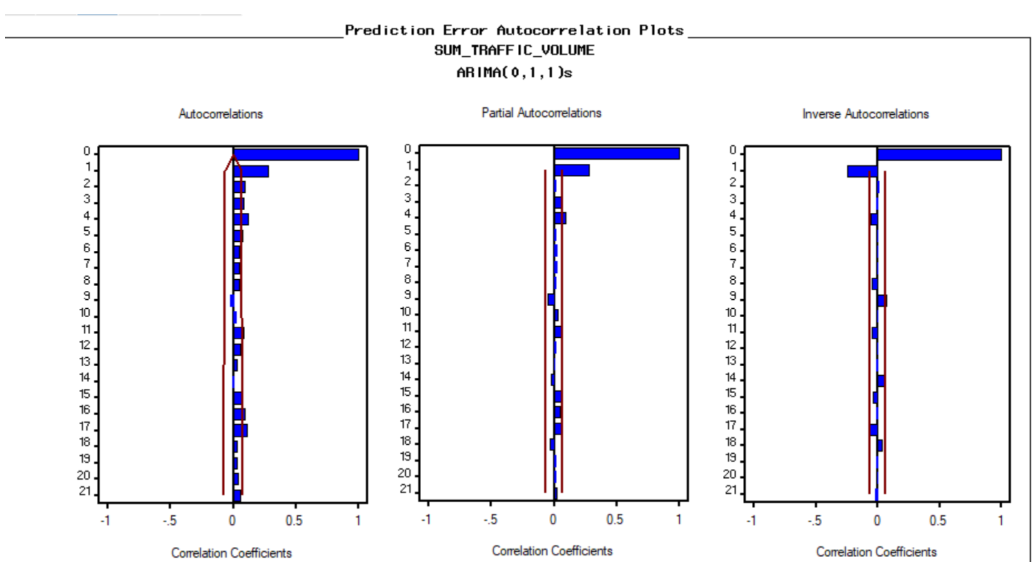
Output 2.10

Output 2.10 shows the behavior of the time series' sample autocorrelation function (ACF), sample partial autocorrelation function (PACF), and sample inverse autocorrelation function (IACF). The sample ACF shows seasonal non-stationary behaviors because the series decays very slowly over seasonal lags such as lag 7, lag 14, and lag 21. Therefore, the first step is to take a seasonal difference. The series' new ACF, PACF, and IACF after this transformation are shown in Output 2.11.



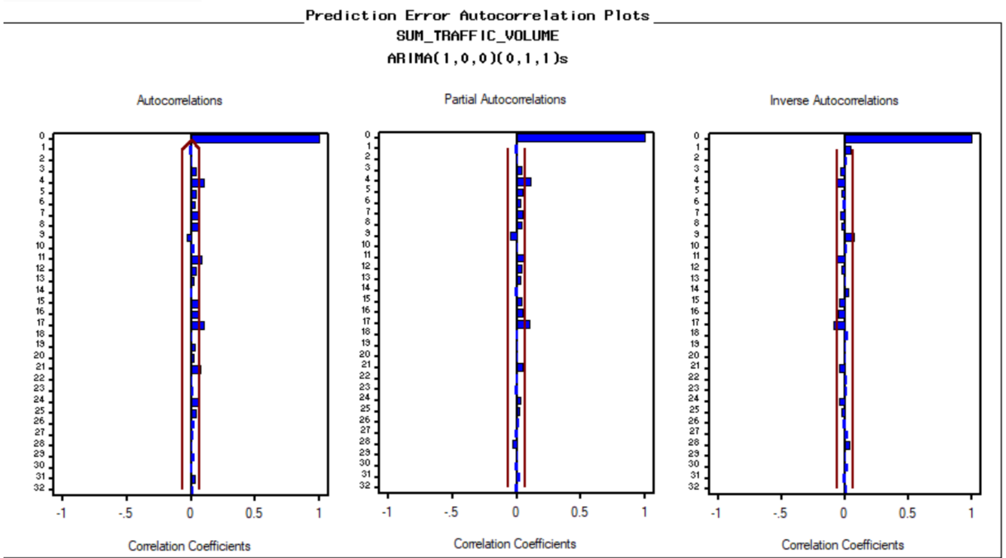
Output 2.11

Now, the series looks seasonal stationary. When we consider only the seasonal lags, the sample ACF is chopped off after lag 7 while the PACF decays exponentially over seasonal lags. Thus, a seasonal MA(1) process with s = 7 is appropriate to start. (s=7 since we have 7 days in a week). We have our first model ARIMA(0,0,0)(0,1,1)s.



Output 2.12

The errors from the ARIMA (0,0,0)(0,1,1)s have the resulting ACF, PACF, and IACF as shown in Output 2.12. Since the ACF of the errors decays over lags and the PACF of the errors basically got chopped off after lag 1, the next model to fit is ARIMA(1,0,0)(0,1,1)s. The resulting prediction error autocorrelation plots from the second model is presented in Output 2.13.



Output 2.13

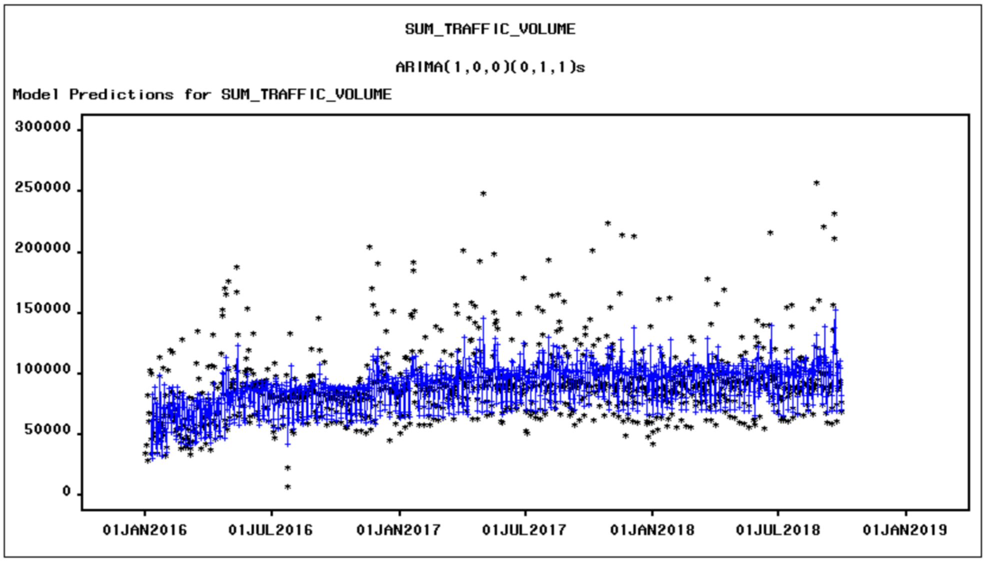
We can see that in the error model, although there are still non-zero autocorrelations at some lags such as lag 4, lag 11, and lag 17, the autocorrelations are not large. Therefore, we can ignore them and consider the resulting series white noise. With the addition of other independent regressors, the model is likely to further improve as the new regressors may explain variation in daily traffic volume that past observations of the time series variable/response variable alone cannot.

Parameter Estimates
SUM_TRAFFIC_VOLUME
ARIMA(1,0,0)(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	194.43804	54.2452	3.5844	0.0005
Seasonal Moving Average, Lag 7	0.96974	0.0130	74.3782	<.0001
Autoregressive, Lag 1	0.29895	0.0326	9.1603	<.0001
Model Variance (sigma squared)	605275484	.	.	.

Output 2.15

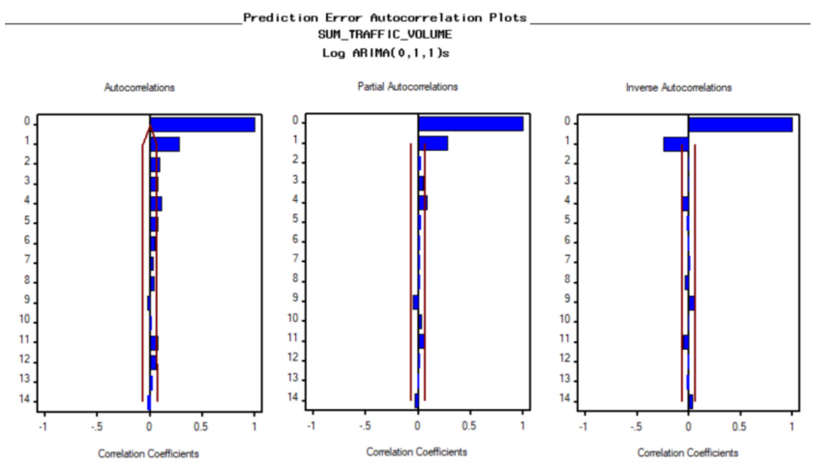
Output 2.15 shows the coefficient estimates as well as the model fit (model variance) of ARIMA(1,0,0)(0,1,1). The measure of model fit, the square root of model variance, is 24602.3. The test RMSE is 27903.0 and the test MAPE is 17.81%. All of the coefficient estimates are significantly different from 0, proving that the model we have chosen is reasonable. We present the plot of actual values versus predicted values from this model in Output 2.16.



Output 2.16

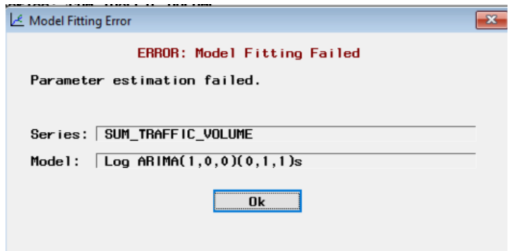
Applying Log Transformation to ARIMA Models: The case for Log ARIMA(1,0,1)x(0,1,1)

As we can see in Output 2.16, after fitting ARIMA(1,0,0) (0,1,1)s to the series, there are still a lot of outliers. Therefore, in the second attempt at ARIMA, we have decided to apply log transformation to the series to alleviate the effect of outliers, after arriving at ARIMA(0,0,0)(0,1,1)s (Output 2.12).



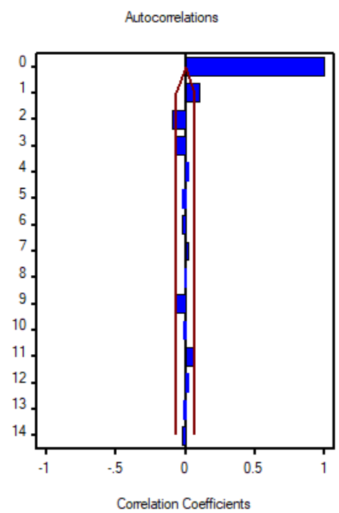
Output 2.17

Output 17 shows that ACF decays over lags and PACF is chopped off after lag1. Therefore, next, we would use an ARMA(1,1). We did not use AR(1) here because the ACF does not decay quickly enough over lags and hence, running Log ARIMA(1,0,0)(0,1,1) would return an error from SAS, as you can see in Output 2.18.



Output 2.18

The prediction error ACF plot of Log ARIMA(1,0,1)x(0,1,1) is presented in Output 2.19 below. The ACF of the error term almost looks like White Noise.

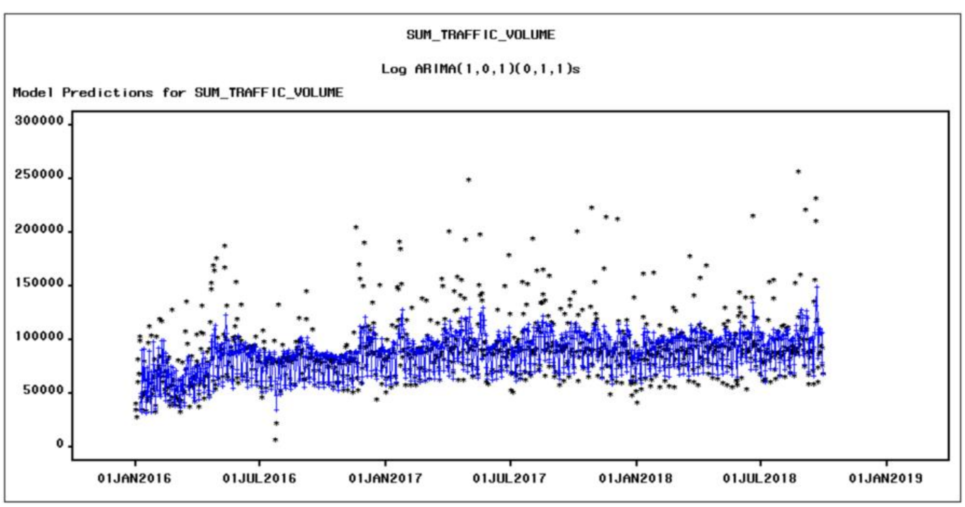


Output 2.19

Parameter Estimates
SUM_TRAFFIC_VOLUME
Log ARIMA(1,0,1)(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.00270	0.000699	3.8558	0.0002
Moving Average, Lag 1	0.68801	0.0479	14.3553	<.0001
Seasonal Moving Average, Lag 7	0.99451	0.0342	29.0380	<.0001
Autoregressive, Lag 1	0.88956	0.0311	28.6212	<.0001
Model Variance (sigma squared)	0.06034	.	.	.

Output 2.20



Output 2.21

Output 2.20 presents the coefficient estimates of the model and Output 2.21 is a time series plot of the actual versus predicted (fitted) values of the series. The intercept and all coefficients are significant. Also in Output 2.20, we learn that the model fit indicator, the square root of model variance, is 0.2.

		Test Error	
	Model Fit (Square Root of Model Variance)	**Validation RMSE**	**Validation MAPE**
ARIMA(1,0,0)(0,1,1)s	24602.3	27903.0	17.811
Log ARIMA(1,0,1)(0,1,1)s	0.2	28418.3	17.650

Table 2.1

2.3. Comparison of Models

		Test Error	
	Model Fit (Square Root of Model Variance)	**Validation RMSE**	**Validation MAPE**
Seasonal Dummies + Holiday Dummy + AR(1)	24601.6	27928.4	14.462
ARIMA(1,0,0)(0,1,1)s	24602.3	27903.0	17.811
Log ARIMA(1,0,1)(0,1,1)s	0.2	28418.3	17.650

Table 2.2

Now, we will compare three univariate models, Seasonal Dummies + Holiday Dummy + AR(1), ARIMA(1,0,0)(0,1,1)s and Log ARIMA(1,0,1)(0,1,1)s in table 2.2. Regarding the performance of two mentioned ARIMA models in section 2.2, we could not compare their model fit performance because one model uses Log transformation and the other does not. Regarding predictive performance, ARIMA(1,0,0)(0,1,1)s performs better according to RMSE, but Log ARIMA(1,0,1)(0,1,1)s performs better according to MAPE.

Moreover, we can see that the deterministic model and ARIMA(1,0,0)(0,1,1) result in similar model fit performance, with Seasonal Dummies + Holiday Dummy + AR(1) performs better by a small margin. Regarding predictive power, although ARIMA(1,0,0)(0,1,1) performs better than the deterministic model by a small margin, the deterministic model has considerably lower Validation MAPE than both ARIMA models.

Therefore, we conclude that **Seasonal Dummies + Holiday Dummy + AR(1)** is the best univariate time series model.

3. Multivariate Time Series Models

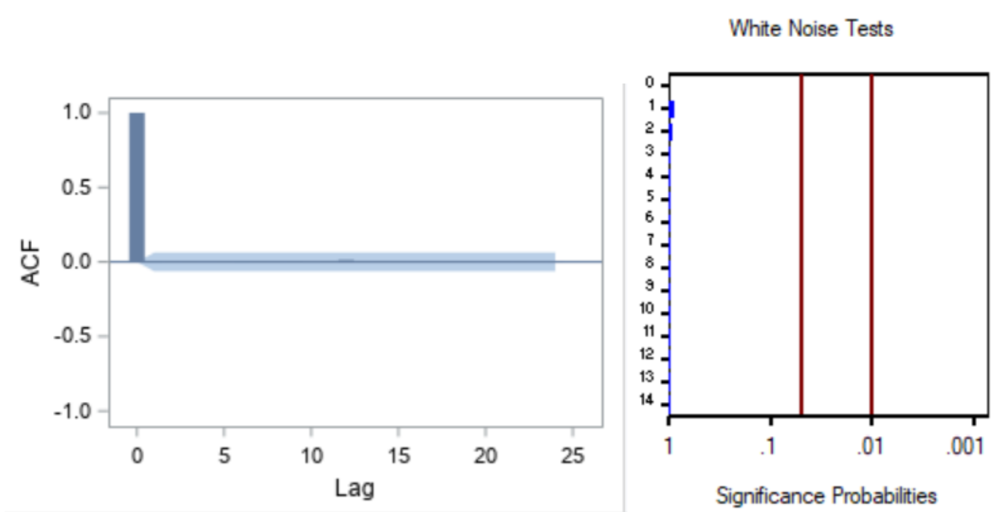
In this section, we will go through the process of developing a transfer function time series model (TF model) step by step. To predict our response variable **sum_traffic_volume**, we will use three independent variables, **average_temp, sum_rain_1h, and average_clouds_all,** as the inputs for our TF model. The description of the variables is provided in section 1 of this report.

The first step is to perform a stationarity check for each of the series (sum_traffic_volume, average_temp, sum_rain_1h, and average_clouds_all). If the series is not stationary, we will perform a simple differencing and(or) seasonal differencing to achieve the stationarity. We also need to ensure that all of the independent variables are white noise. We will pre-whiten the variables if necessary using the ARMA process. Next, we will identify and estimate individual TF models to present the relationship between the response variable, sum_traffic_volume, and each of the independent variables based on the respective cross-correlation function (CCF) . Then, we will check the adequacy of the individual models, i.e. to check the CCF of the input and the residuals. Next, we will estimate each multiple-input TF model and check its adequacy as well. In the last step, we will check the ACF of the residuals of the multiple-input TF model to see whether a TF-noise model is necessary.

a. Sum_traffic_volume

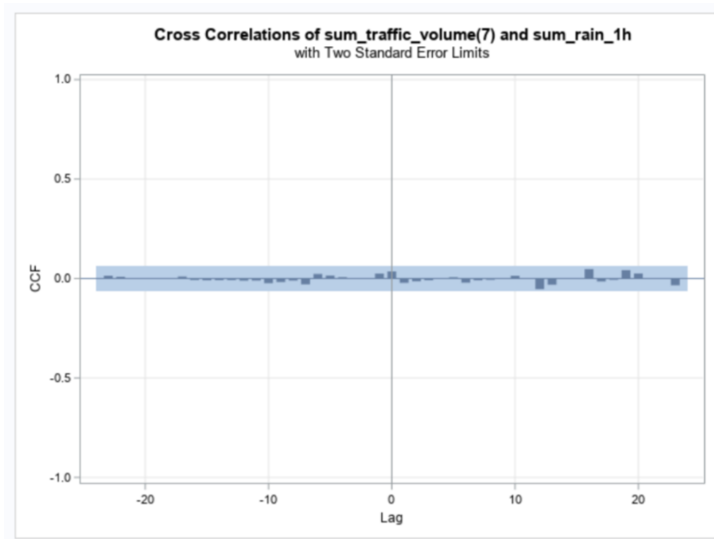
The series looks non-seasonal and seasonal stationary after the first seasonal differencing. Please refer to Output 2.11 under section *2.2. ARIMA Models* to view the series' ACF after the transformation.

b. Sum_rain_1h:



Output 3.1

Based on the ACF and white noise tests results in output 3.8, the sum_rain_1h variable is both stationary and white noise. Thus, no preprocessing and pre-whitening is required.

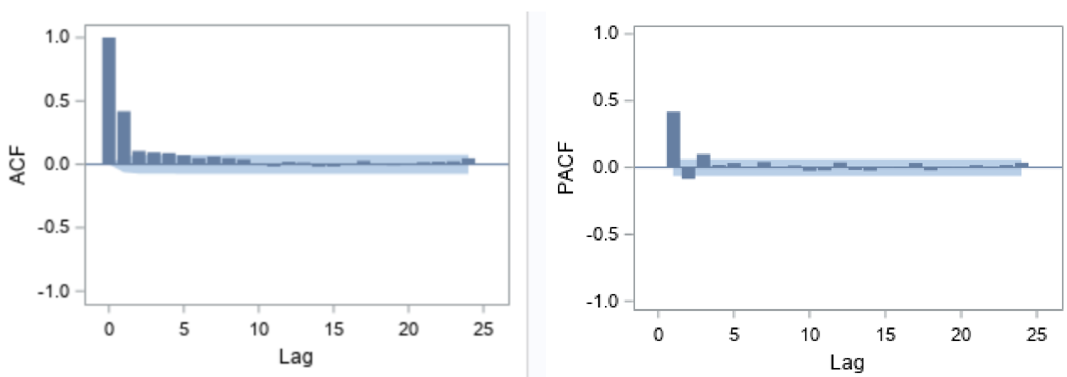


Output 3.2

Output 3.9 shows that there exists no lag where the cross-correlation between sum_traffic_volume(7) and sum_rain_1h variable is significantly different from 0. Therefore, we will not include the sum_rain_1h variable in our multivariate TS model.

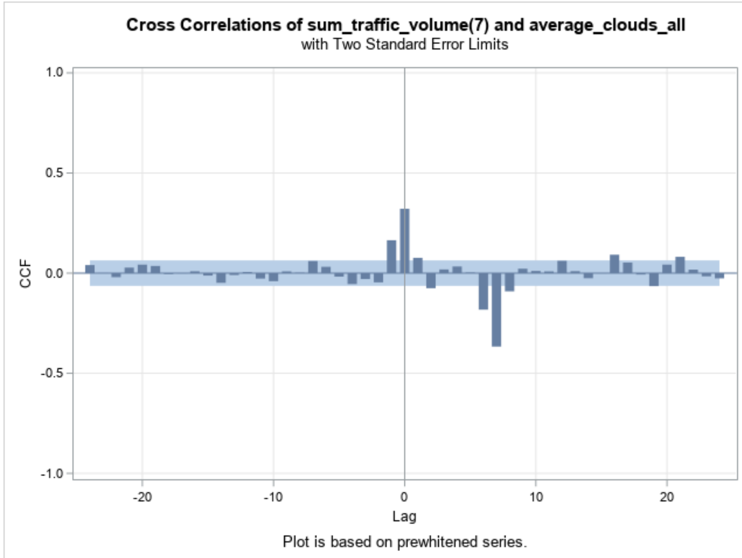
c. Average_clouds_all:

Output 3.10 presents the ACF plot of average_clouds_all variable. We treat the series as a stationary but not WN series. The ACF and the PACF are both decaying and thus, we use ARMA(1,1) to prewhite the series.



Output 3.3

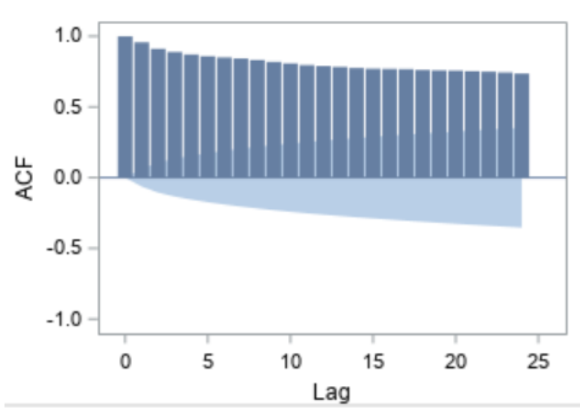
To check whether the variable is an appropriate variable for the TF model, we look at the CCF between the response variable and the average_clouds_all variable, in Output 3.11.



Output 3.4

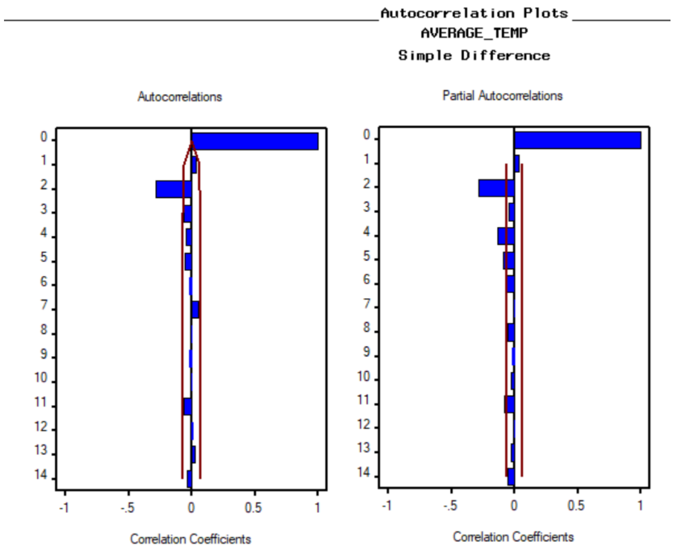
The negative spike indicates that the relationship between the sum_traffic_volume and the average_clouds_all is unidirectional. Therefore, we will not include the average_clouds_all as a TF input.

d. Average_temp



Output 3.5

From the ACF plot in Output 3.5, we can see that average_temp is not stationary. It behaves similarly to that of a random walk model: slowly decays over lags. Thus, we take the first difference of the series.

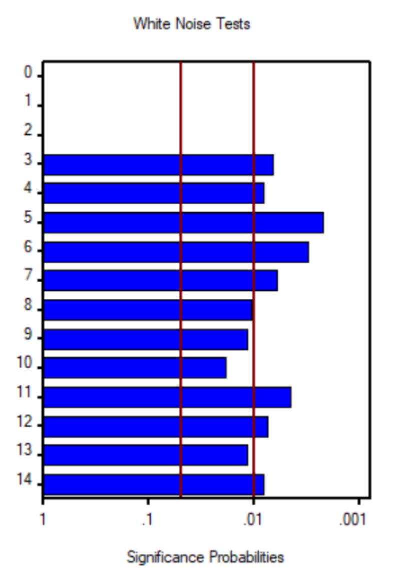


Output 3.6

Output 3.6 presents the ACF of average_temp variable after the first difference. It now looks stationary but not white noise as the autocorrelation at lag 2 is significantly different from 0. This observation is strengthened with Output 3.7 below. Therefore, we will need to pre-whiten the first difference of average_temp using an MA(2) process before fitting it into a TF model.

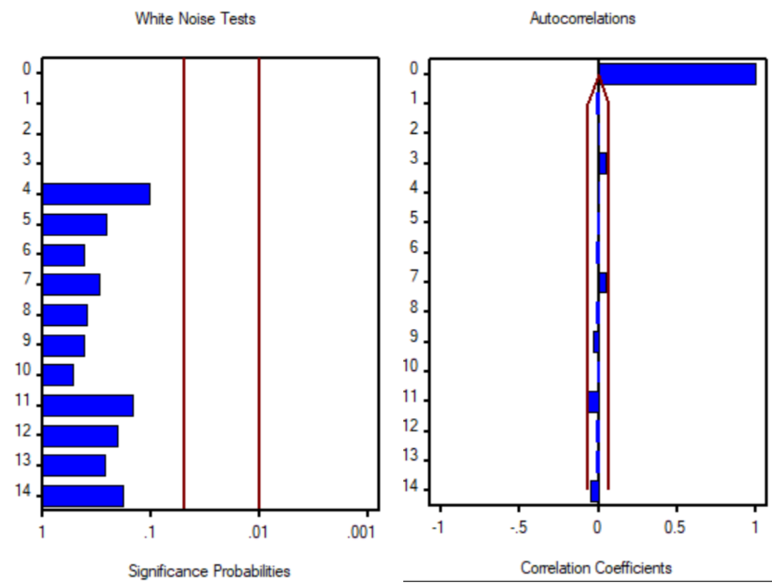
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	89.16	6	<.0001	0.034	-0.280	-0.063	-0.045	-0.052	-0.014
12	95.97	12	<.0001	0.053	-0.003	-0.019	-0.005	-0.058	0.011
18	104.08	18	<.0001	0.025	-0.037	-0.053	-0.006	0.050	-0.023
24	107.06	24	<.0001	-0.008	0.033	-0.012	0.023	0.033	-0.001

Output 3.7

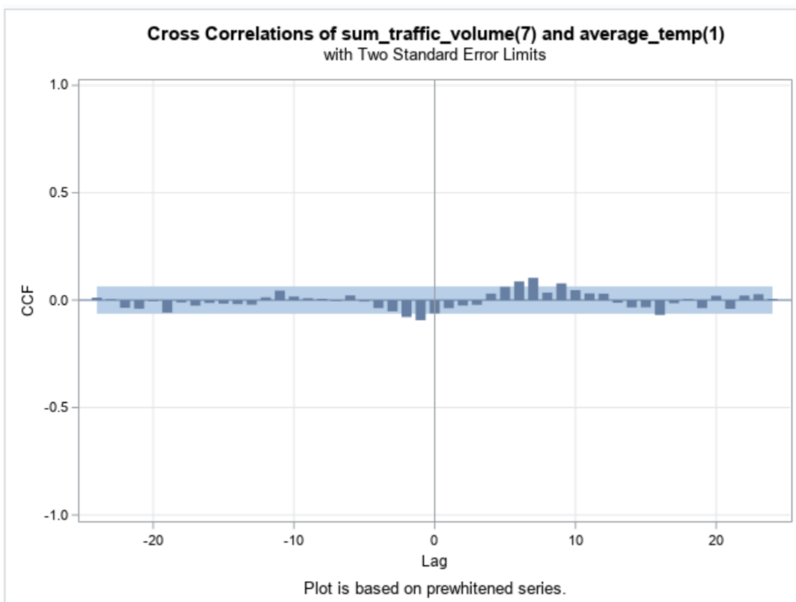


Output 3.8

However, after ARIMA(0,1,2), the independent variable, average_temp, is still not WN, as seen in output 3.8. Therefore, we will add AR(1) to the pre-whitening process. The result of the ARIMA(1,1,2) suggests that the series is now a WN TS, output 3.9.



Output 3.9



Output 3.10 (b=6, s=1, r=2)

The CCF suggests that it is appropriate to add this variable into the TF model. The CCF becomes non-zero from lag 6 thus b = 6. The CCF is decaying from lag 6 thus b = 6 and s = 1. The CCF displays sinusoidal behavior, therefore r = 2.

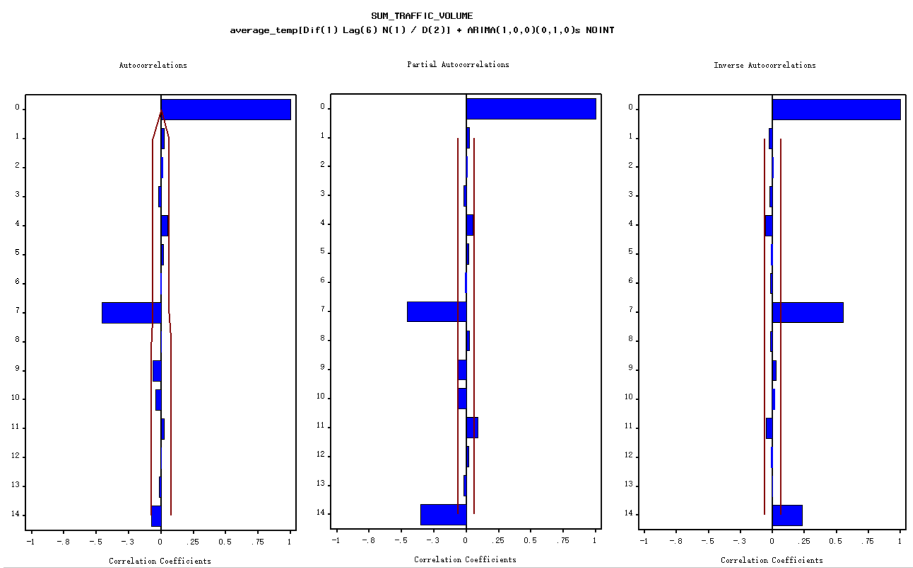
After estimating the individual TF model for the sum_traffic_volume and average_temp variables, we check the adequacy of the model, i.e., the cross correlation between the input and the residuals. Output

3.11 shows that the input average_temp and the residuals are not cross-correlated. Therefore, it is appropriate to include this variable in our multiple-input model.

Crosscorrelation Check of Residuals with Input average_temp									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	4.63	2	0.0986	0.027	0.038	-0.026	0.023	0.029	0.022
11	7.77	8	0.4566	0.038	0.001	-0.015	0.006	-0.022	0.032
17	13.70	14	0.4723	0.055	-0.018	0.029	-0.041	0.000	0.014
23	19.99	20	0.4585	-0.030	-0.055	0.004	-0.005	0.038	-0.030
29	20.66	26	0.7593	-0.001	-0.019	-0.014	-0.004	0.010	-0.006
35	21.53	32	0.9194	-0.012	-0.013	0.021	-0.008	0.002	-0.008
41	30.92	38	0.7857	-0.017	0.078	0.002	0.055	0.008	0.003
47	39.08	44	0.6820	0.022	0.019	-0.071	0.008	-0.048	0.003

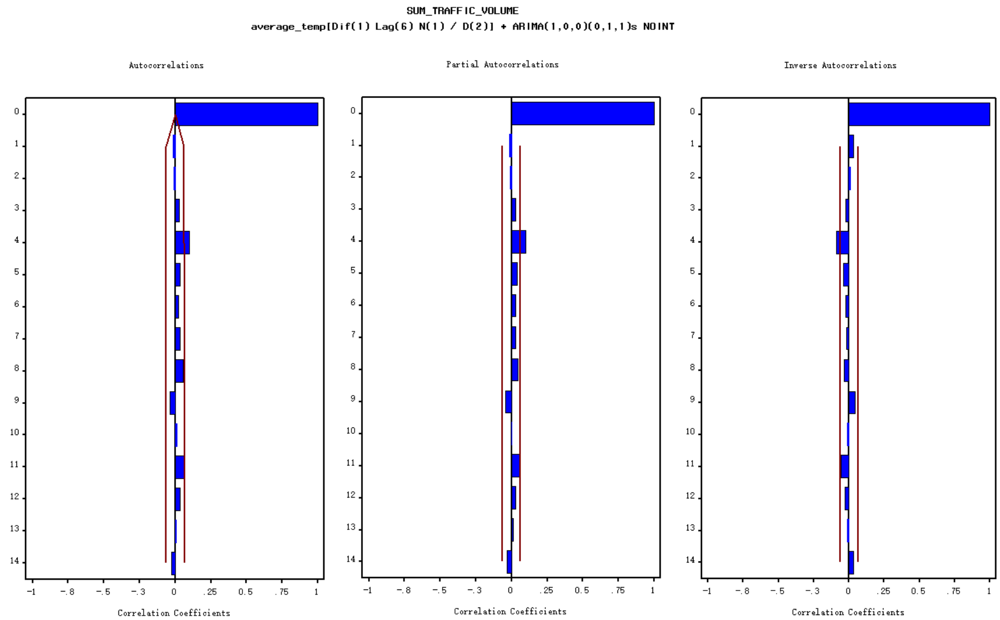
Output 3.11

The final TF model has one variable as input which is the average_temp variable.



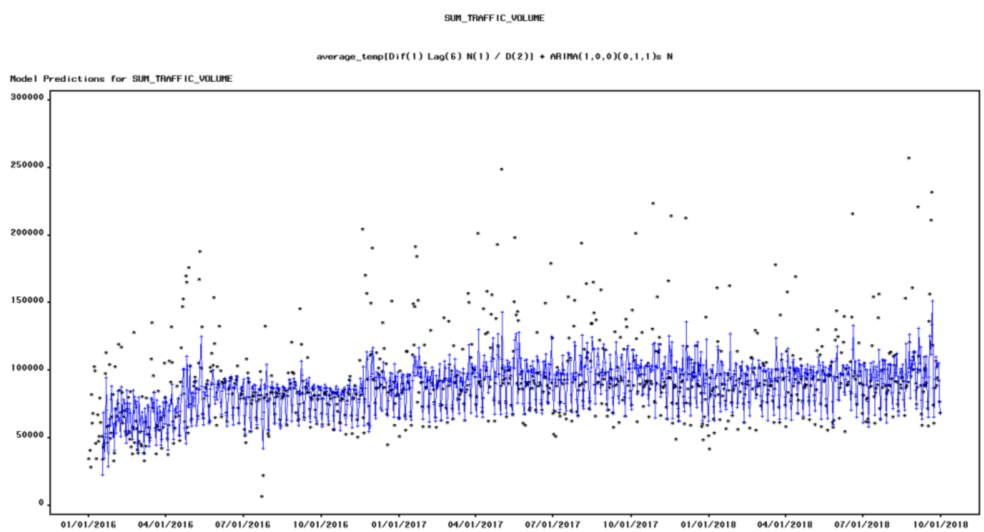
Output 3.12

According to the Output 3.12 the ACF of the TF model residual suggests that the model is not WN yet. Therefore, we then included an error model into the TF model. The ACF and PACF have a significant spike at lag 7. Since this is a daily time series data, the spikes indicate a seasonal signal. Thus, we included a SARIMA(1,0,0)(0,1,1) for the error model.



Output 3.13

In the Output 3.13 the TF error model has a WN residual and this is the appropriate model.



Output 3.14

SUM_TRAFFIC_VOLUME
average_temp[Dif(1) Lag(6) N(1) / D(2)] + ARIMA(1,0,0)(0,1,1)s NOINT

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Seasonal Moving Average, Lag 7	0.93833	0.0138	68.1770	<.0001
Autoregressive, Lag 1	0.29278	0.0331	8.8493	<.0001
AVERAGE_TEMP[Dif(1) Lag(6) N(1) / D(2)]	35.04705	240.4743	0.1457	0.8843
AVERAGE_TEMP[Dif(1) Lag(6) N(1) / D(2)]	-565.85299	238.4952	-2.3726	0.0190
AVERAGE_TEMP[Dif(1) Lag(6) N(1) / D(2)]	0.12549	0.3913	0.3207	0.7489
AVERAGE_TEMP[Dif(1) Lag(6) N(1) / D(2)]	0.30881	0.3693	0.8363	0.4044
Model Variance (sigma squared)	613524137	.	.	.

Output 3.15

SUM_TRAFFIC_VOLUME
average_temp[Dif(1) Lag(6) N(1) / D(2)] + ARIMA(1,0,0)(0,1,1)s NOINT

Statistic of Fit	Value
Mean Square Error	801503570
Root Mean Square Error	28310.8
Mean Absolute Percent Error	15.97252
Mean Absolute Error	17656.1

Output 3.16

Output 3.15 shows the coefficient estimates of the model and output 3.16 shows its predictive performance.

4. Conclusion

In table 4.1 below, we present the final model comparison between the best univariate models and the best multivariate TF model. Based on both the model fit and two test error metrics, we can conclude that the best model for traffic volume is **Seasonal Dummies + Holiday Dummy + AR(1).**
According to output 2.9a, when it is Saturday and not a holiday, the traffic volume will be 75475 (model intercept value). If a day is a Sunday or a holiday, then the traffic volume will be lower than the referenced Saturday-not-a-holiday. We were originally surprised by the negative coefficient of the Holiday dummy, but when we thought about it more, traffic tends to increase some days before the holiday, and people usually don't travel as much on the holiday itself. Moreover, the traffic volumes on Monday, Tuesday, Wednesday, Thursday, and Friday are all higher than Saturday, with Friday having the highest increase in traffic volume (coefficient estimate of seasonal dummies). Therefore, if the government wants to increase toll on a specific day of the week, it should be Friday.

		Test Error	
	Model Fit (Square Root of Model Variance)	**Validation RMSE**	**Validation MAPE**
Seasonal Dummies + Holiday Dummy + AR(1)	24601.6	27928.4	14.462
ARIMA(1,0,0)(0,1,1)s	24602.3	27903.0	17.811
TF b = 6, s = 1, r = 2, SARIMA(1,0,0)(0,1,1)	24769.4	28310.8	15.973

Table 4.1