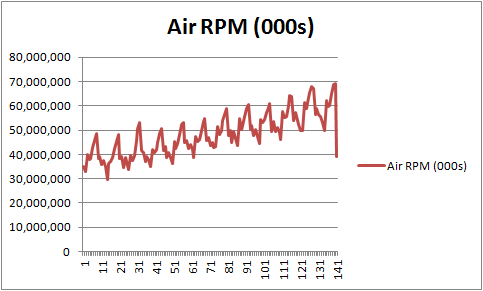
1. The plot of the pre-event Air time series is shown below

**ASSIGNMENT 3**

**SIDDHARTH BHATTACHARYA (61510787)**



The components that appear from the plot are

* There is a definite trend upward as time goes on.
* There is seasonality with window size of approximately 12 months.
* There is a level at each period.

1. The best method to forecast the outcome of this series would be to use linear regression with trend .Here it is important to realize that in the seasonality adjusted plot the seasonality has been removed. Thus with the seasonality removed there is only trend which is there in the plot. Hence the optimum choice would be go for **“Linear Regression Model with Trend”.**
2. A linear regression trend for the series that would produce a seasonally adjusted series is constructed, first we introduce a variable called window which takes dummy values and are 11 in mumber(for window size of 12,this is done through Transform option).Next we introduce make the series exponential to account for multiplicative seasonality. We take log(Air Miles) with the predictors being the dummy variables apart from time t.

d) On running the model the following model is obtained.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input variables** | **Coefficient** | **Std. Error** | **p-value** | **SS** |
| Constant term | 7.56674194 | 0.00873465 | 0 | 10160.18359 |
| t | 0.00117037 | 0.00004461 | 0 | 0.56961107 |
| Window\_1 | -0.02530117 | 0.01078839 | 0.02025076 | 0.03052144 |
| Window\_2 | -0.05170499 | 0.01078756 | 0.0000037 | 0.08985394 |
| Window\_3 | 0.03540329 | 0.01078692 | 0.00126685 | 0.00055283 |
| Window\_4 | 0.01267737 | 0.01078646 | 0.24162996 | 0.0045143 |
| Window\_5 | 0.02464273 | 0.01097492 | 0.02612309 | 0.00069369 |
| Window\_6 | 0.05942592 | 0.01097374 | 0.00000018 | 0.0121524 |
| Window\_7 | 0.08656401 | 0.01097275 | 0 | 0.058454 |
| Window\_8 | 0.09498747 | 0.01097193 | 0 | 0.10985249 |
| Window\_9 | -0.00688651 | 0.01097129 | 0.53111136 | 0.0001494 |
| Window\_10 | 0.01210087 | 0.01097084 | 0.27169237 | 0.00486198 |
| Window\_11 | -0.02144595 | 0.01097057 | 0.05235294 | 0.00321945 |

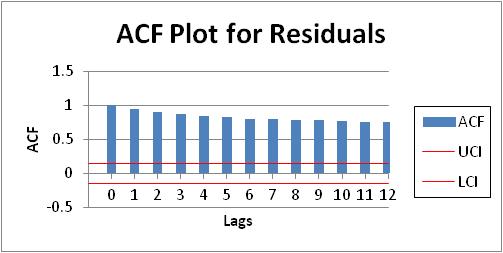
The 11 coefficients correspond to the 11 dummy variables for a seasonal window of 12.

The graph for the predicted values over the training prevent data shows the following (this is only on the pre-event data)

The residuals show the following trend over the same prevents data.

1. From the table we see that for September and October the p values are significantly higher than the threshold (of 0.05) compared to the other months, this indicates that the values are insignificant. It indicates that the actual airline revenue passenger miles in these months are less (air travel might be less).
2. The Actual value for January 1990 was 35.153577 million while the predicted value was 43.7billion, so this gives a residual value of 8.59 billion. Here the values had to obtained by raising the values back to exponent.

e) i) The ACF plot as shown below



This plot tells us that there is a strong positive correlation. This tells us that the forecasted values need to be adjusted for the positive autocorrelation which is occurring and this can be directly fit into the regression model for more accurate forecast.

The positive autocorrelation which is occurring can be directly adjusted into the original regression model to adjust and get more accurate forecasts; this can also be done by running a second level forecasting model on the residual time series. Now when we are forecasting say a k step ahead we need to first forecast the k-step ahead forecast of forecast error, using the AR model and then improve the initial k step ahead forecast of the series by adjusting it according it back to the original forecast ie F\*t+k = Ft+k + et+k.

f)

The model for air including trend and additive seasonality is shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input variables** | **Coefficient** | **Std. Error** | **p-value** | **SS** |
| Constant term | 33433400 | 541313.875 | 0 | 2.9405E+17 |
| t | 171750.8594 | 3571.803711 | 0 | 5.7078E+15 |
| Window\_1 | -2144908 | 673519.4375 | 0.00185413 | 2.70197E+14 |
| Window\_2 | -4667999 | 673434.1875 | 0 | 7.20212E+14 |
| Window\_3 | 4000806 | 673367.875 | 0.00000003 | 3.71704E+11 |
| Window\_4 | 1603923 | 673320.5625 | 0.01881221 | 6.01263E+13 |
| Window\_5 | 2865584 | 673292.125 | 0.00004195 | 1.87281E+13 |
| Window\_6 | 6651891 | 673282.625 | 0 | 7.35829E+13 |
| Window\_7 | 9914961 | 673292.125 | 0 | 4.86564E+14 |
| Window\_8 | 11188560 | 673320.5625 | 0 | 1.015E+15 |
| Window\_9 | 882650.8125 | 673367.875 | 0.19247061 | 3.11546E+12 |
| Window\_10 | 2382147 | 673434.1875 | 0.00057886 | 7.50891E+13 |
| Window\_11 | -1609106 | 673519.4375 | 0.0184746 | 1.35531E+13 |

For Auto the regression model is similar and is shown below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input variables** | **Coefficient** | **Std. Error** | **p-value** | **SS** |
| Constant term | 167.2879334 | 0.86860037 | 0 | 7603284 |
| t | 0.41108462 | 0.00443602 | 0 | 70515.67969 |
| Window\_1 | -12.94057369 | 1.07283032 | 0 | 6805.202148 |
| Window\_2 | -20.81555367 | 1.07274783 | 0 | 15091.49023 |
| Window\_3 | 7.79115248 | 1.07268357 | 0 | 376.9252625 |
| Window\_4 | 7.36318016 | 1.0726378 | 0 | 557.2058716 |
| Window\_5 | 19.8869648 | 1.09137928 | 0 | 578.0084229 |
| Window\_6 | 18.67402649 | 1.0912621 | 0 | 523.7649536 |
| Window\_7 | 26.22024918 | 1.09116292 | 0 | 3368.08252 |
| Window\_8 | 26.51265526 | 1.09108174 | 0 | 5230.867188 |
| Window\_9 | 6.66849566 | 1.09101868 | 0 | 58.44437408 |
| Window\_10 | 14.18206406 | 1.0909735 | 0 | 2047.27417 |
| Window\_11 | -1.25753701 | 1.09094644 | 0.25076243 | 11.06961155 |

For Rail we need to fit a quadratic equation and so we first need to add the t^2 parameter and then construct the regression after that. This is shown below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input variables** | **Coefficient** | **Std. Error** | **p-value** | **SS** |
| Constant term | 577495500 | 11480320 | 0 | 3.36157E+19 |
| t | -2053446 | 243484.6094 | 0 | 1.61978E+17 |
| t^2 | 8251.173828 | 1661.142212 | 0.0000021 | 2.29935E+16 |
| Window\_1 | -70724300 | 12173060 | 0 | 6.7814E+16 |
| Window\_2 | -92201100 | 12171260 | 0 | 1.32805E+17 |
| Window\_3 | 3198767 | 12169890 | 0.79309773 | 3.05972E+15 |
| Window\_4 | 2990214 | 12168950 | 0.80629188 | 3.95394E+15 |
| Window\_5 | 13319910 | 12168440 | 0.27575153 | 1.10948E+15 |
| Window\_6 | 37602100 | 12168350 | 0.00245856 | 2.74949E+15 |
| Window\_7 | 87244210 | 12168690 | 0 | 6.33952E+16 |
| Window\_8 | 95082780 | 12169460 | 0 | 1.17427E+17 |
| Window\_9 | -25999800 | 12170670 | 0.03457595 | 1.53346E+15 |
| Window\_10 | -12527860 | 12421830 | 0.31511495 | 1.06791E+12 |
| Window\_11 | -25817130 | 12421390 | 0.03968201 | 3.6658E+15 |

**i)**

Now we need to fit when the entire data is taken

**For rail** the plot of predicted values before was (pre event data)

And after words was following

**For Air the 2 plots were as follows**

Pre-event

PostEvent

**For Auto** the plots look as follows

PreEvent

PostEvent

**ii** So looking at the plots we can safely say that Air transportation decreased after the terrorist attack. The Auto transportation increased where as the Rail transportation seems to have increased slightly compared to the predicted values. It may have been that many of the passengers who earlier preferred Air travel switched to Road or Rail for safety issues.

**Solution 4**

The plot for the differenced series is given below

A random walk is a series where if we look at the differenced series we find that future values cannot be predicted on basis of lagged values, this is an indication of autocorrelation existing in the data and hence not a random series(more precisely a random walk).Also if the coefficient of AR(1) is 1 the series is a random walk(this is because if we transpose the time related terms on LHS, we are left with only error terms on RHS).So now if we examine the options given in the question we find that option 3 (AR(1) constant coefficient for the closing price) and option 6(AR(1) constant coefficient for the differenced series) are related to the coefficient of the constant term not the slope and hence is not a test of random walk. All the others are a test of Random walk.

c)

On recreating the AR(1) model we find the following ACF plot

A ARIMA model is now run with get the following output

|  |  |  |  |
| --- | --- | --- | --- |
| **ARIMA** | **Coeff** | **StErr** | **p-value** |
| Const. term | 2.30948114 | 0.06041662 | 0 |
| AR1 | 0.95588744 | 0.01867162 | 0 |

From this table we can clearly see that the coefficient of AR(1) is 1 and is significant at this level(given by the p-value),so it is an indication of a random walk.(for getting to this result an ARIMA model is run with Autoregressive as 1,Diffrence and Moving Average set to 0).

d) When a series is following a random walk there is no useful information that can be predicted. So using Forecasting techniques is not helpful. The series is a random series and future values donot depend on past values, so all the three statements given are actually true.

SOLUTION Chapter 9 Problem 1

From the data set we see that for every hour, there are 4 equispaced intervals and it repeats every hour ,hence this can be captured through 3 dummy variables(n-1 variables required: period2,period3,period4).The response variable here is the demand every 15 minutes.

The output after the regression is run is summarized below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Input variables** | **Coefficient** | | | **Std. Error** | **p-value** | **SS** |
| Constant term | 25.75528717 | | | 1.29424393 | 0 | 886147.9375 |
| Period\_2 | -0.21148036 | | | 1.83033729 | 0.90806252 | 53.01758957 |
| Period\_3 | 0.71440995 | | | 1.83172333 | 0.69668871 | 106.3536911 |
| Period\_4 | 0.03865239 | | | 1.83172333 | 0.98317307 | 0.24688406 |
| Residual df | | 1318 |
| R-squared | | 0.000218379 |
| Std. Dev. estimate | | 23.54670334 |
| Residual SS | | 730761.5 |

|  |  |  |
| --- | --- | --- |
| **Total sum of squared errors** | **RMS Error** | **Average Error** |
| 730761.4779 | 23.51105379 | -1.60197E-07 |

**Actual/Predicted Plot**

**Residual plot.**