Name: Nirupa Dissanayake

Time Series Analysis Student Project

Time Series Project: France Tourism Industry

Introduction

France is the most popular tourist destination in the world which makes about 30% of its GDP. France is located in the heart of the Western Europe. The location, historic heritage and the diversified natural environment makes the country a hot spot for tourism industry. The purpose of this project is to study the tourism performance patterns over past couple of years.

Data

Quarterly tourist visit counts (from 1993-2015) from all over the world and the percentage change compared to same period in previous year were used as the data for this project. Data was collected from http://data.worldbank.org and can be found in "Data" tab in the attached excel sheet.

Methodology

Stationary Time series

A time series is stationary if there is no systematic change in mean or variance over the periods, in other words no periodic trend of the data.

Autocorrelation

Autocorrelation is the correlation of time series with itself. A positive autocorrelation means that an increase in time series is followed by another increase and vice versa.

Differencing

Computing the differencing between consecutive data points is a way to make non stationary time series stationary.

Seasonal differencing

Seasonal differencing is the difference between one period and the corresponding period in previous year. Seasonal differencing is used to convert non stationary time series with seasonality to stationary time series.

ARIMA model

ARIMA (p,d,q) models are a method of forecasting by converting non-stationary time series to stationary time series using differencing where,

- **p** - Order of the autoregressive terms.
- **d** - Degree of differences need to be stationarized.
- **q** - Order of lagged forecast errors in the prediction equation.

Data Analysis

Original Data Analysis

Summary of the data can be found in Figure-1. There is an overall increasing trend of the tourist visits over the period of 1993 to 2015. As expected, data shows clear seasonal pattern between quarters. Graph shows higher volumes in second and third quarter each year compare to other two. Therefore it can be concluded that the data is not stationary.

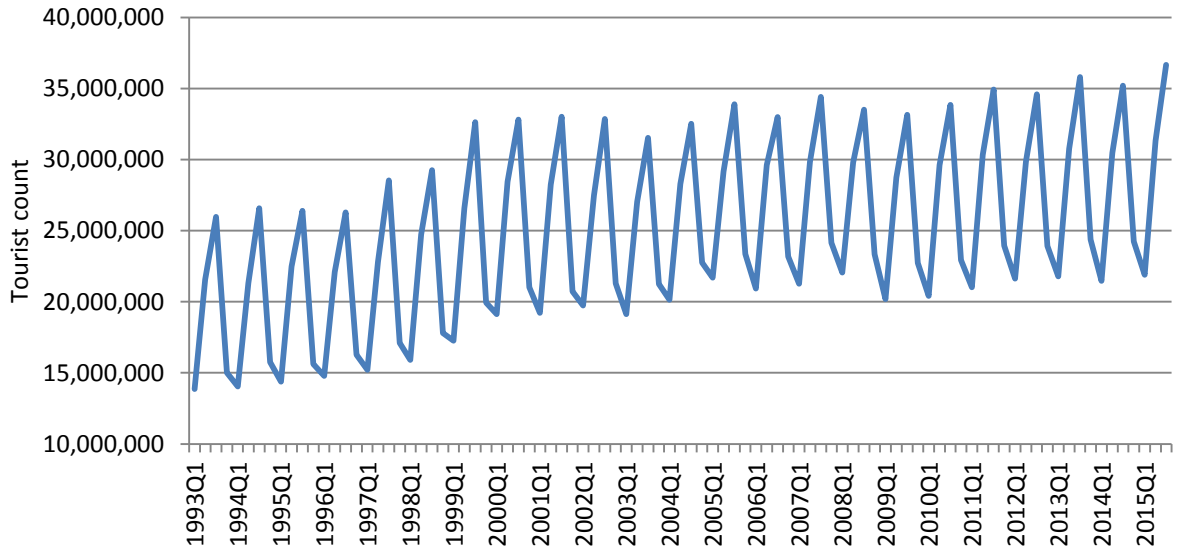

Figure 1 - Quarterly Tourist Count

Figure-2 shows the percentage change of tourist counts from the same period of previous year. There is a huge percentage drop in 2009 compared to 2008 due to economic crisis and unfavorable currency rates that happened in 2009.

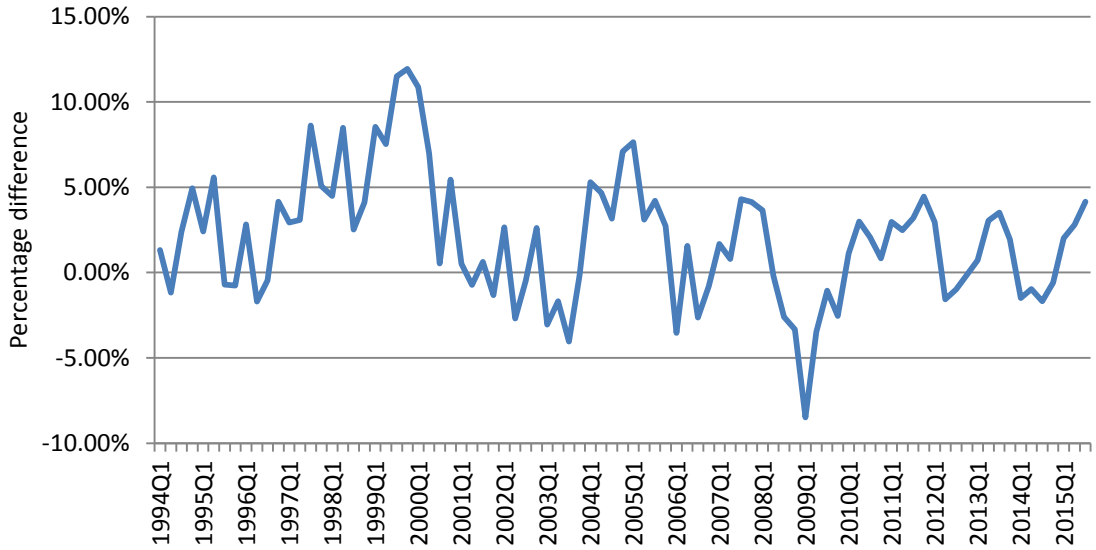

Figure 2 - Percentage Change of tourist count compared to previous Year

First Differece Analysis and Autocorrelation

To make the data stationary first difference of the quarterly data was taken as shown in Figure-3. Now the general upward trend has disappeared but still seasonality can be shown. Autocorrelation of the first difference was calculated and it's found that the series oscillates around zero as shown in figure-4 (Please refer to the attached spreadsheet tab "data"). By looking at both graphs we can conclude that the most of the seasonal difference is gone but not all.

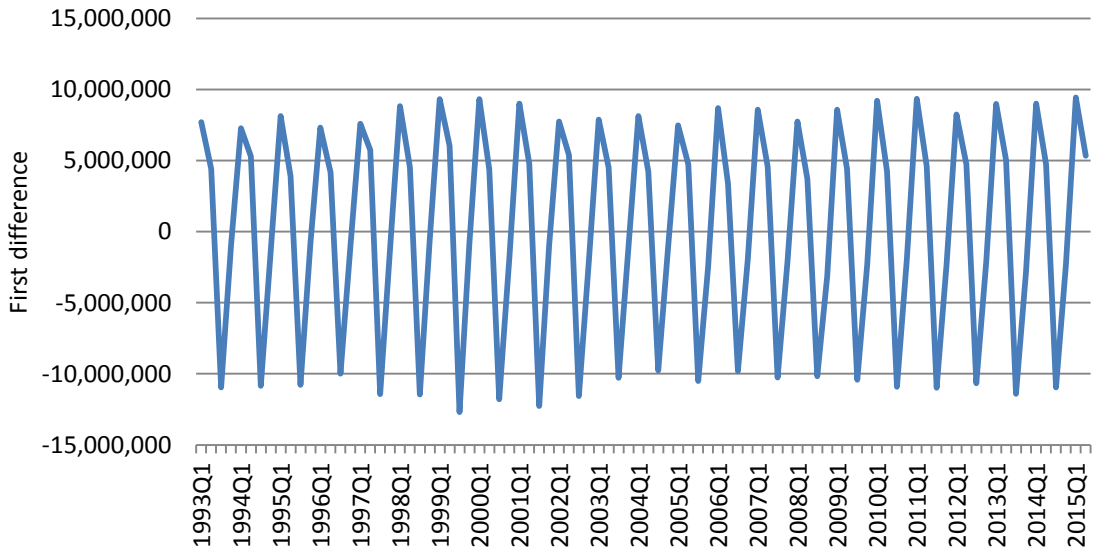

Figure 3 -First Difference of quarterly tourist count in France

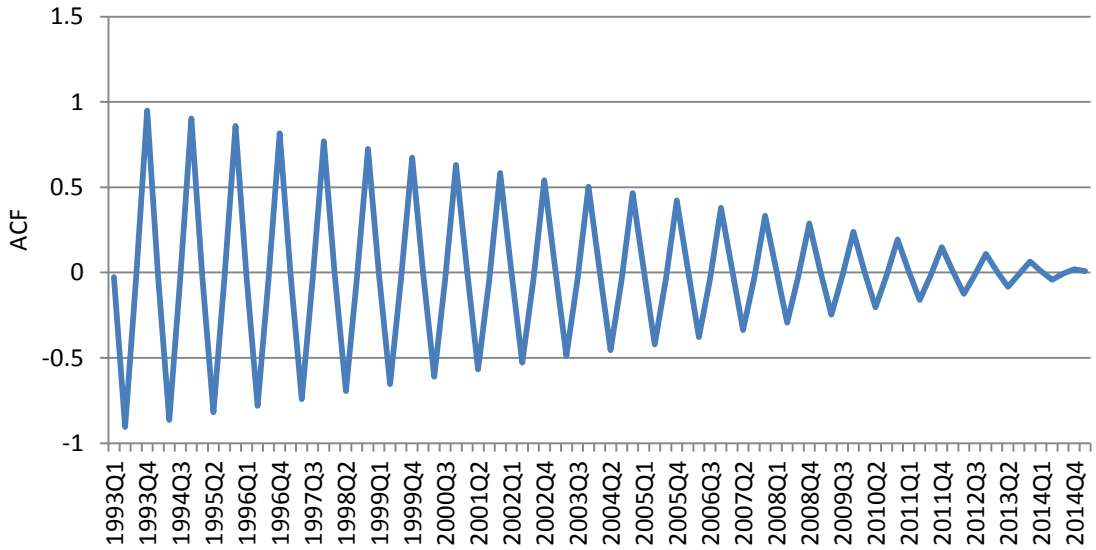

Figure 4 – Autocorrelation of first difference of quarterly tourist count in France

First and Seasonal Difference

Figure-5 and Figure-6 show a very little autocorrelation remains in the series after the two changes. Due to the increasing trend and seasonal effect presented in the original data, it can be concluded that the time series is multiplicative seasonal ARIMA (0, 1, 1) x $(0, 1, 1)_4$ forecasted by, $Y_t = Y_{t-1}+Y_{t-4} - Y_{t-5}+ e_t - \theta e_{t-1} - \emptyset e_{t-4}+ \theta\emptyset e_{t-5}$

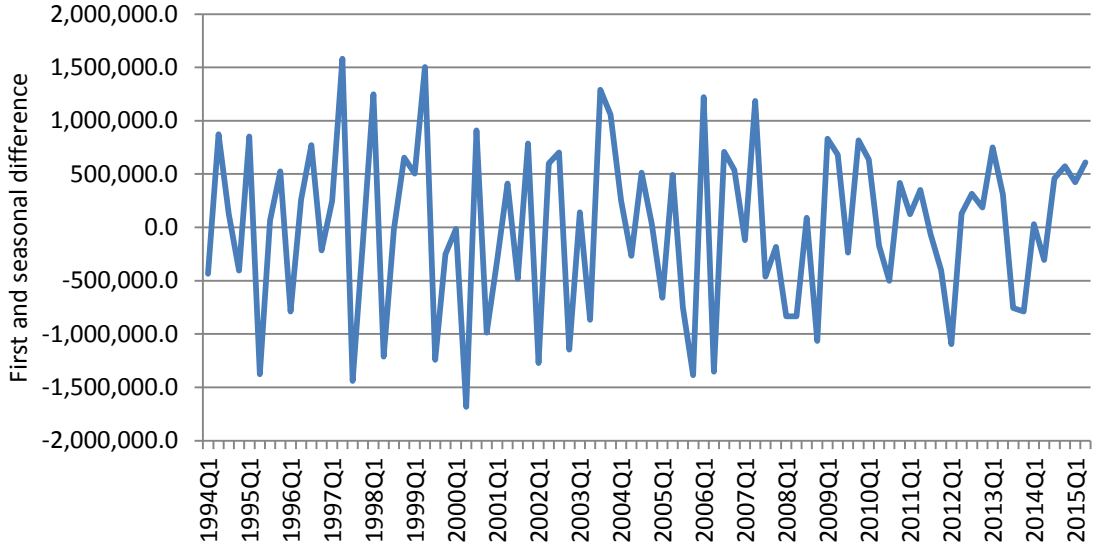

Figure 5 – First and seasonal difference of tourist count

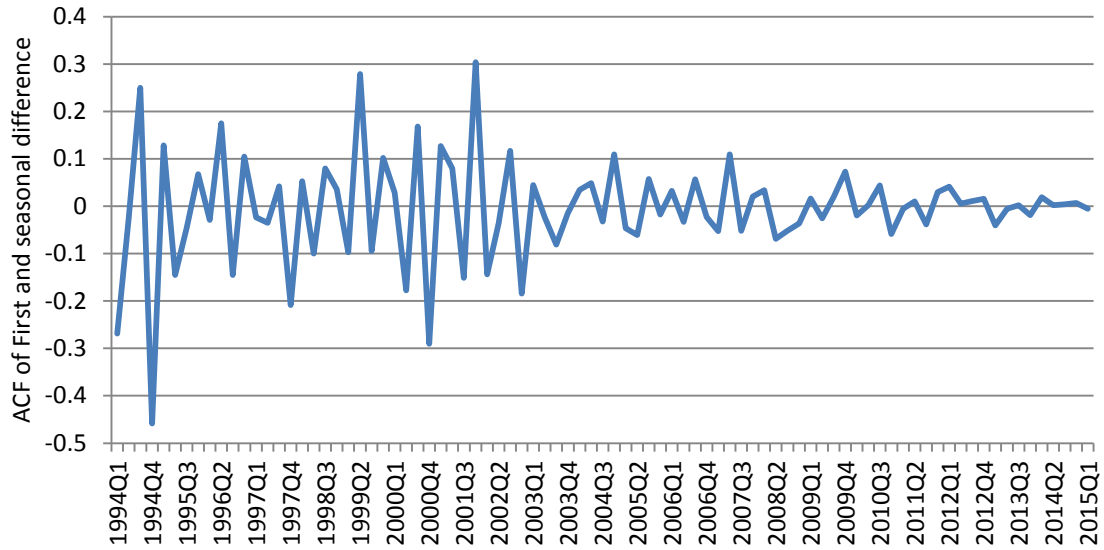

Figure 6 – Autocorrelation of First and Seasonal difference

Model Fitting

Having specified a seasonal model for the France tourist data, we can estimate the parameters for the model. Parameters were estimated using XLSTAT (please see the attached spreadsheet tab "ARIMA"). The forecasted time series equation is $Y_t = Y_{t-1}+Y_{t-4} - Y_{t-5}+ e_t + .279e_{t-1} + .521e_{t-4}+ .145e_{t-5}$ where Moving Average parameter is -0.279 and the seasonal moving average parameter is -0.521.

Diagnostic Checking

Figure-7 shows the estimated values vs the actual values and it can be seen that the estimated values fit very closely to the actual values.

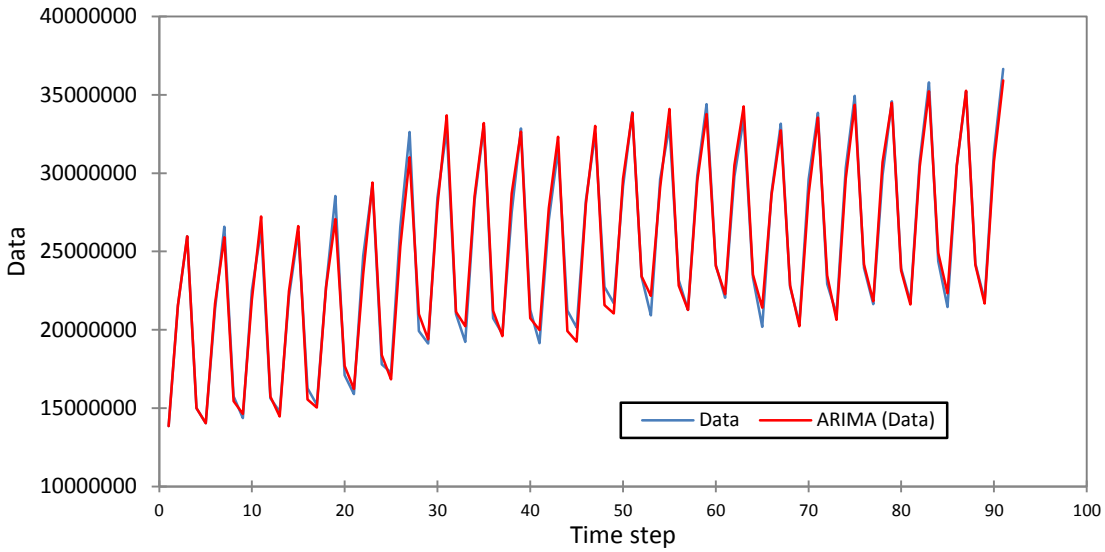

Figure 7 – Estimated vs Actual values

To check the reasonability, time series plot of residuals were graphed in Figure-8. This figure gives a plot of standardized residuals. Residuals were calculated using XLSTAT and can be found in the attached spreadsheet in tab "ARIMA". The series does not show any considerable irregularities other than few outliers.

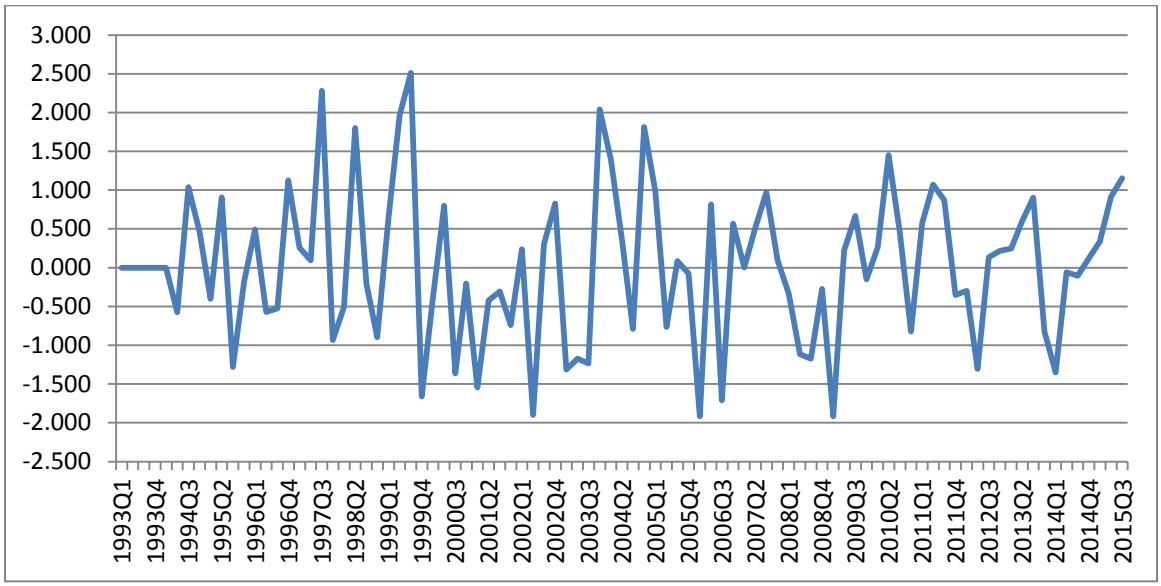

Figure 8 - Residuals from ARIMA

To look further, quantile - quantile plot was graphed to investigate normality as seen in Figure-9. There are few outliers in extremes but no major irregularities present. This is sufficient to conclude that the residuals are normally distributed. To summarize, it can be said that the above forecasted equation is a good fit for the data given for tourist count in France from 1993 to 2015.

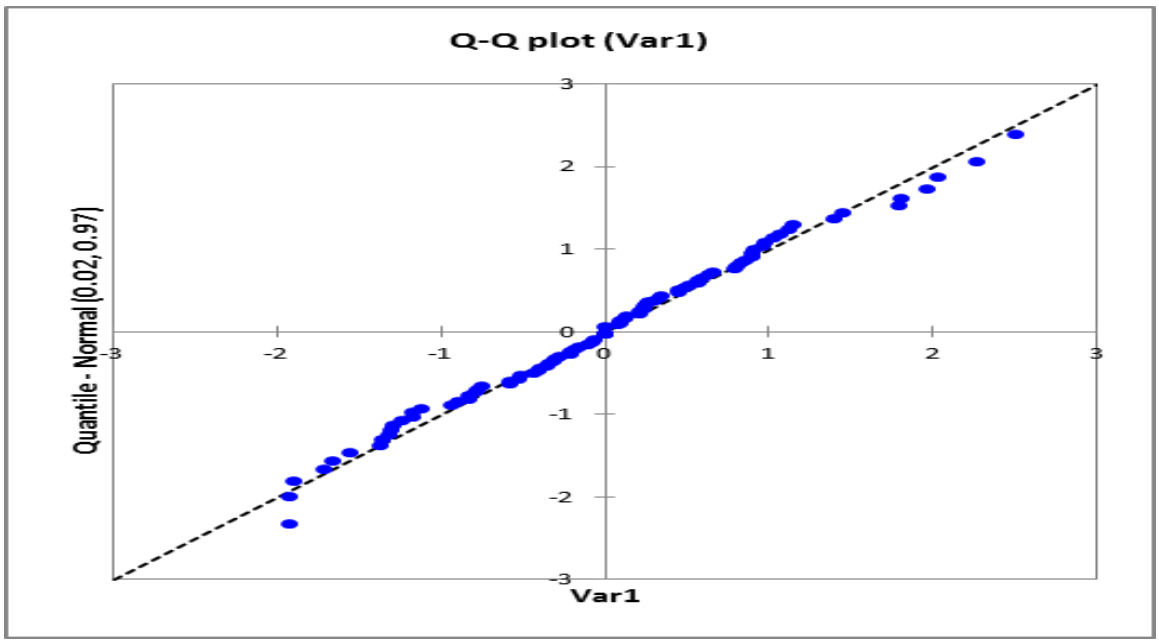

Figure 9 - Q-Q plot of standardized residuals