Applied Statistical Methods – Time Series (Summer 2014)
Student Project

Joseph Daryl Abesamis
jgabesamis@gmail.com

Air Traffic at Airports in the UK

Introduction

Airports have always fascinated me. It amazes me how their system can handle so many passengers at any given time, all of whom are taking specific flights, departure times, connections and destinations.

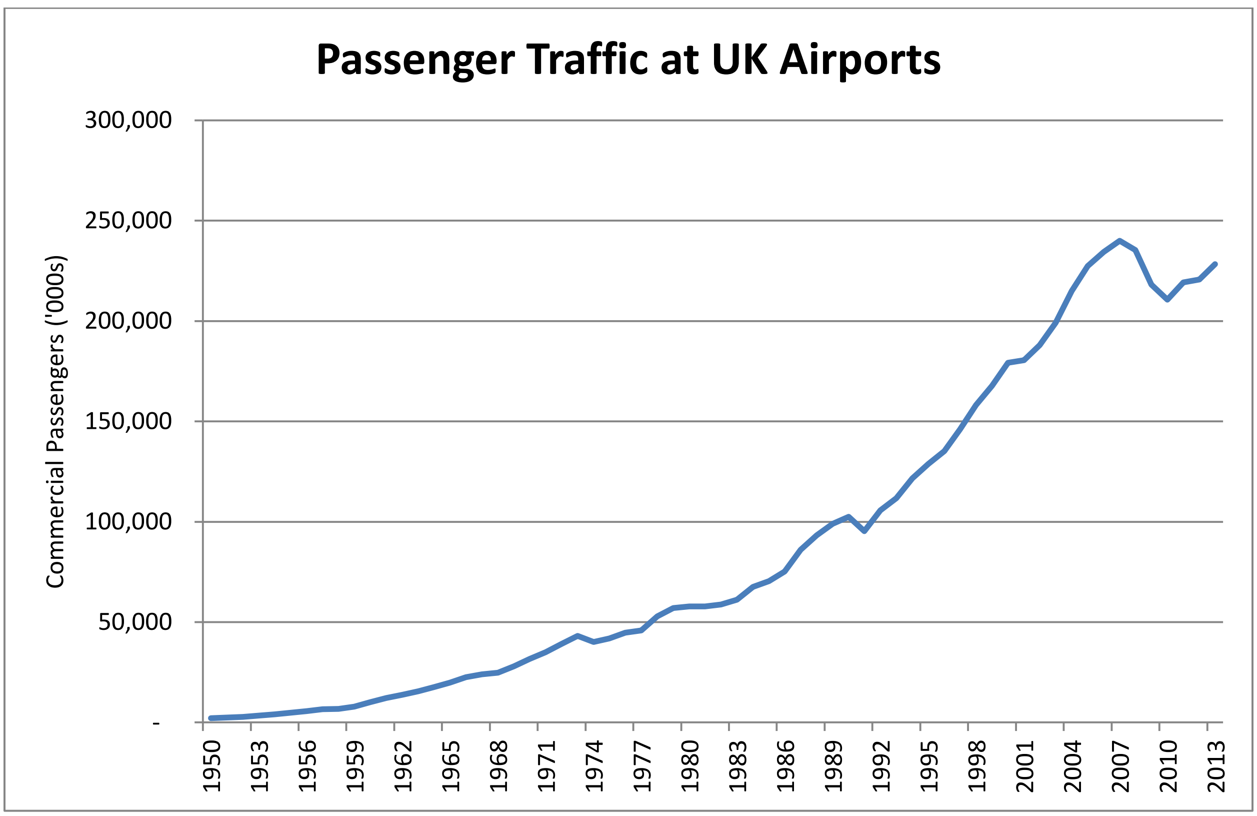
I've decided to look at airport traffic for the past 60 years. I've decided to focus on a single country – the United Kingdom. I've obtained data from the UK's Department for Transport website:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/339556/avi0101.xls

I will attempt to find a model to predict the future traffic for the UK's airports.

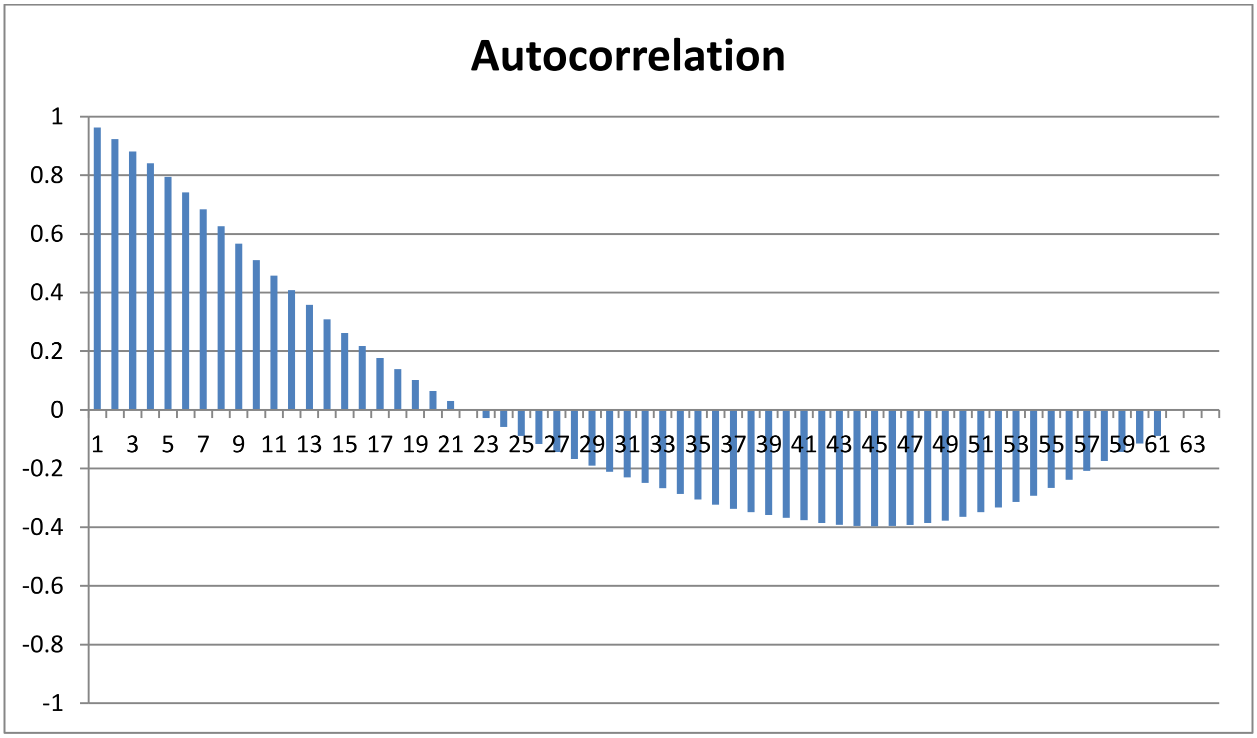
Data

The graph of the total terminal passengers (arrival and departures) beginning 1950 are shown below. It is apparent that the number of passengers is increasing steadily.



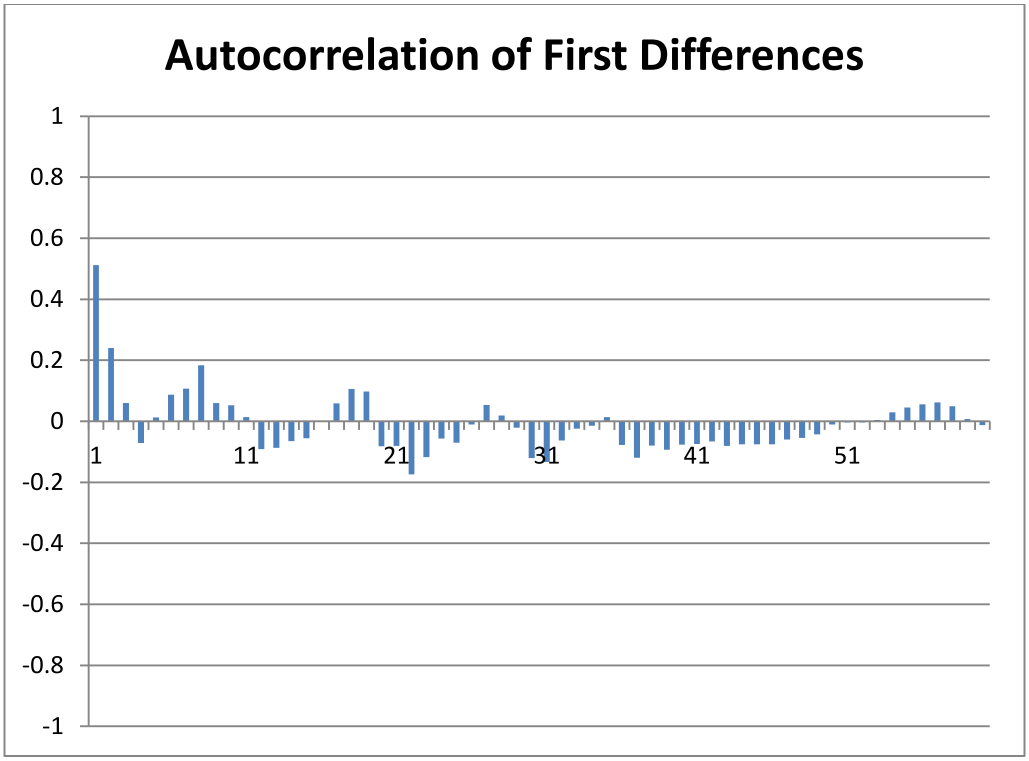
There were few years where there is a decline in the number of passengers. There may be a number of factors which cause a decline in the number of commercial air travel passengers. The oil crisis in 1974, the economic slowdown from 1990 to 1992, and the 2008 financial crisis could have contributed to these declines.

It is necessary to determine stationarity, which would allow the process to be modeled using fixed coefficients estimated from prior data. A generally increasing trend indicates non-stationarity. To verify this, the sample autocorrelation of the data is examined.



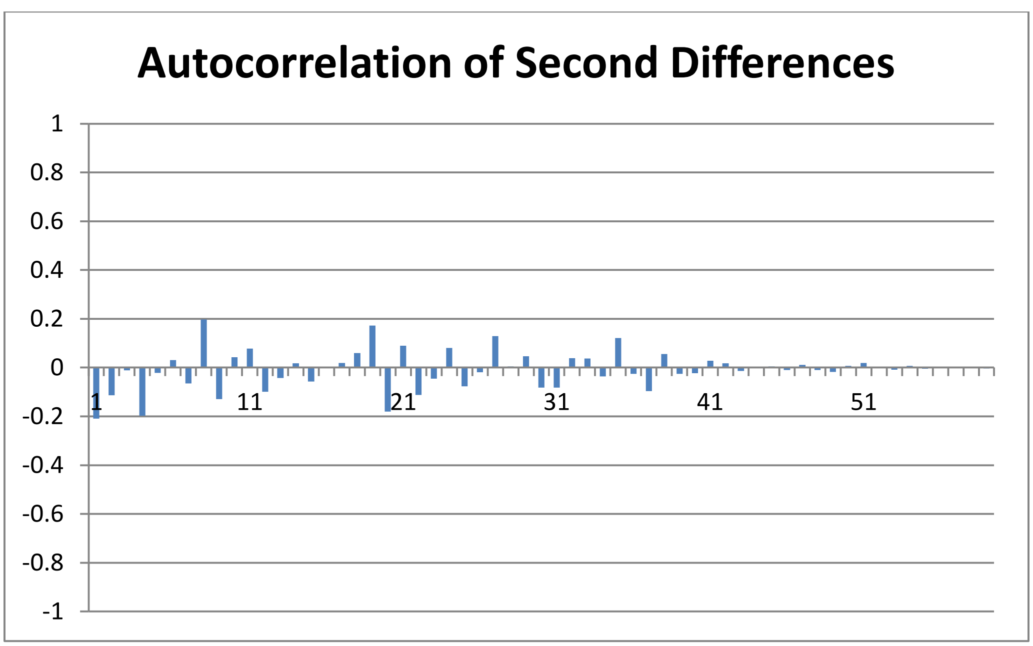
While the sample correlations tend to zero, it does so slowly. This indicates non-stationarity.

The autocorrelations of the first difference are also examined to check for stationarity.



The plots of the sample autocorrelation of the first difference appear close to zero and oscillate around zero. This implies stationarity.

As an additional check, the sample autocorrelation of the second difference of the time series was also examined. It inferred stationarity, as well (as expected.)



Model Specification

Regression was done on the first differences of the original time series. Three models were tested:

$$\text{ARIMA}(1,1,0)\text{:}\quad W_t = \delta + \phi W_{t-1} + \epsilon_t$$

$$\text{ARIMA}(2,1,0)\text{:}\quad W_t = \delta + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \epsilon_t$$

$$\text{ARIMA}(3,1,0)\text{:}\quad W_t = \delta + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_3 W_{t-3} + \epsilon_t$$

Note: W_t is the first difference of the time series, i.e. $W_t = Y_t - Y_{t-1}$.

After running the regression, we get the following coefficients.

	δ	ϕ_1	ϕ_2	ϕ_3
ARIMA(1,1,0)	1820.634515	0.517284126		
ARIMA(2,1,0)	1920.328506	0.532973092	-0.035982174	
ARIMA(3,1,0)	2056.518742	0.52731057	-0.003829406	-0.059855908

For each model, we look at the Durbin-Watson statistic, the Box-Pierce Q statistic and the Adjusted R^2 to determine which model to use.

Statistic	ARIMA(1,1,0)	ARIMA(2,1,0)	ARIMA(3,1,0)
Durbin-Watson	1.94611	1.98304	1.99845
Box-Pierce Q	15.50061	15.57268	14.51395
Chi-squared (10%)	72.15984	71.03971	69.91851
Adjusted R^2	0.25431	0.23893	0.22449

The Durbin-Watson statistic is used to determine whether or not the residuals or correlated. The ideal value of this statistic is 2, meaning no serial correlation of residuals. All three models exhibit a Durbin-Watson statistic that is close to 2, with ARIMA(3,1,0) having the closest value.

The Box-Pierce Q statistic indicates whether or not the residuals form a White Noise process. The null hypothesis is that the residuals form a White Noise process. This hypothesis cannot be rejected if the Q statistic is lower than the critical Chi-squared value. At 10% significance, all three models exhibit residuals possibly forming a White Noise process.

The Adjusted R^2 statistic determines how well-fit the model is, and is a measure of how well the regression line models actual data. The ideal value of this statistic is 1. Of the three models, the ARIMA(1,1,0) model produced the Adjusted R^2 closest to 1. However, the value is still very low indicating that it is a poor fit. This indicates that all three models are poor fits to the time series.

Conclusion

While the residuals indicate a good model, given the low Adjusted R^2 of each model, the actual process of the underlying Air Traffic Passengers have not been determined. It is highly possible that determining the process would require greater sophistication than the methods done in this project.

Of the three models, should exactly one *must* be selected, the $\text{ARIMA}(1,1,0)$ model would be the best to work with because it has the "best fit" among the three. Is also has the lowest p parameter for the $\text{AR}(1)$ process of the first difference.

The $\text{ARIMA}(1,1,0)$ model is described by the following equation of the first difference. The graph comparing actual and modeled Air Traffic Passenger follows.

$$W_t = 1820.64 + 0.52W_{t-1} + \epsilon_t$$