

UK TRAVEL AND TOURISM FORECASTS

2016



MS Business Analytics and Project Management

**OPIM 5671 – Data Mining and Business Intelligence**

**Project Report - Summer 2016**

Team 9

Sujith Vallepalli

Pranitha Buddiga

Shreya Chandra

Aswajit Mohapatra

Table of Contents

[Executive Summary 3](#_Toc456218802)

[Methodology 5](#_Toc456218803)

[Data Preparation 5](#_Toc456218804)

[Data Exploration: 6](#_Toc456218805)

[Approach: 8](#_Toc456218806)

[Modelling: 8](#_Toc456218807)

[Forecasting 9](#_Toc456218808)

[Model Equations: 14](#_Toc456218809)

[Important Observations from Forecasts: 14](#_Toc456218810)

[Conclusion/Recommendation: 16](#_Toc456218811)

[Appendix 17](#_Toc456218812)

[Visualization Using Tableau 121](#_Toc456218813)

[SAS Codes & Data Sets: 124](#_Toc456218814)

# Executive Summary

**Business Context:**

UK is the eighth largest international tourism destination ranked by both number of visitors and visitor expenditure (Source: United Nations World Tourism Organization data). It accounts for 3.5% of global international tourism receipts. Since 2010 tourism has been the fastest growing sector in the UK in employment terms generating one-third of all jobs.

The big picture - the tourism economy: delivering jobs and growth:



Source: Tourism: jobs and growth. Deloitte November 2013



Source: Tourism: jobs and growth. Deloitte November 2013

Seasonality and trend inspected across breadth of measures in both inbound and outbound tourism of UK. We have forecasted measures like:

1. Monthly forecasts on number of visits and spending for inbound and outbound UK tourism from 1980 to 2015.
2. Yearly region wise number of visits to and from UK from 1980 to 2015 and forecasted till 2018. We accounted for visits to UK from North America, Europe and other countries broadly and vice versa for outbound travel from UK. We also forecasted the spending for both inbound and outbound
3. Purpose of visits (business, holiday, friends visited and others) to and from UK over a period of 1980 to 2014 and the forecasted till 2018. We also forecasteddifferent transportation modes like Air, ship and channel tunnels. For channels the time series analysis was done from 1994 to 2014 as it was opened in 1994 and for the rest it was from 1980 onwards.

For all the forecasted models, we talk about the number of visitsin thousands and the spending in million pounds.

We considered trend and seasonality while forecasting all the above parameters.

We have also done visualization of the different measures additionally in Tableau to derive key business insights. These tableau presentation and dashboards were prepared from the forecasted data derived from different final models.

Please refer appendix section for detailed forecasts and models.

**Benefits of forecast numbers:**

Government's tourism strategy (Competing against a tough international environment, Britain may add value to the tourism industry by building our competitive tourism offer and driving economic growth and jobs across the nations. Building internal capacity for transportation, hotels and traffic system at the right time and at right place and right price should be the top priority.

**Benefits to other industries:**

* Govt. Sector – It can accelerate many new marketing initiatives to brand UK as a tourist destination of choice. Focus - Sustainable tourism.
* Marketing & Media industry: Budget allocation for marketing campaigns across different business sectors not limited to hotel, food and transportation industry.
* Transport Industry – Although number of visits forecasted are almost same, Air transportation has 18,000 million spending compared to much above other modes. Marketing budgets can be designed accordingly.
* Food Industry &Hotel Industry - New generation of cloud base systems disrupting the landscape. Exciting opportunity ahead.

Business drives and government initiatives can take advantage of the forecasted measures:

* Employment: One-third jobs in UK were from the tourism industry from 2010 to 2015. One-third of the tourism job are in accommodation, one-third in catering and food, 15% in retailing & attraction and the rest in transport.
* Smart city example Bristol - Government-backed UK Smart City initiative - installed more than 200 beacons around the city, which use Bluetooth connectivity to send and receive real time tourist data.
* Driver less cars in Milton Keynes: self-driving pods & visitors avail facility through smartphone app.

# Methodology

## Data Preparation

The dataset that we used for preliminary forecasting is monthly visits and earnings data for both inbound and outbound for 1980-2015. Source of our data is from Office of National Statistics, which maintains national archives for UK gov. We have 7 excel files with different parameters with respect to Inbound and Outbound tourism over time. But these files have data in different timeframes. that is, yearly data from 1980-2015, monthly data from 1980 to 2015, and quarterly data from 2004 to 2014. Since our data is a tourism data of UK, we realized multiple economic variables which are not part of the original data may bring more insights. So we have done our research to get the monthly average data for socio-economic variables like oil prices per barrel, inflation in UK, temperatures in UK, GBP value per USD. We appended these new variables data to most of the original files to bring more meaning to the original time series data.

**Variables and description**

|  |  |
| --- | --- |
| Variables | Description |
| Date | Date variable (YYMON) |
| Gbpusd | GBP to USD conversion rate |
| Oilusd | Oil prices in USD |
| Oilgbp | Oil prices in GBP |
| Inflation | Inflation rate in UK |
| Visitsuk | Number of visits to UK (in thousands) |
| visitsabroad | Number of visits abroad (in thousands) |
| Earnings | Money spent in UK by inbound visitors (in Million £) |
| expenditure | Money spent abroad by outbound visitors of UK (in Million £) |
| temp\_in\_C | Monthly averaged temperature(ᵒC) |

Target variables identified are “**Total visits**” and “**Spending**” for inbound and outbound tourism

Before performing time series modeling techniques, the team checked for various types of data anomalies like incorrect datatypes, inconsistences in the data, and missing observations.

*Modification of “Date” field –* When the data was uploaded for modeling, SAS failed to identify the field “Date” as a valid field. “Date” was of the format “mmm-yy” which was not picked up by SAS as a valid date element and it gave us errors.

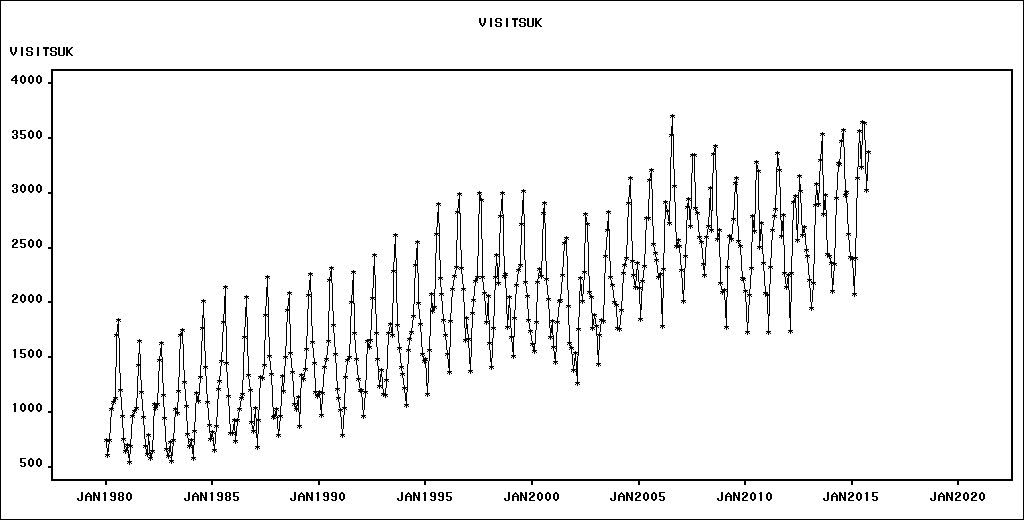
So we used SAS code to import the raw files in a way that SAS can recognize all the variables required correctly.

We checked for missing field values to ensure that the data isn’t patchy and also to ensure that the data that is collected over time is continuous

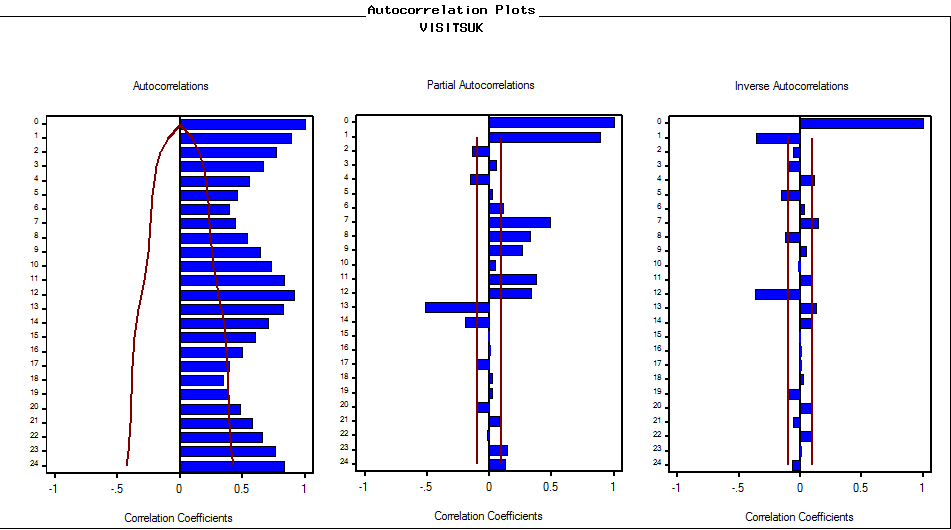
Our data is now read into Time Series Forecasting System and its ready for forecasting (Refer to appendix for SAS code: ***import.sas***)

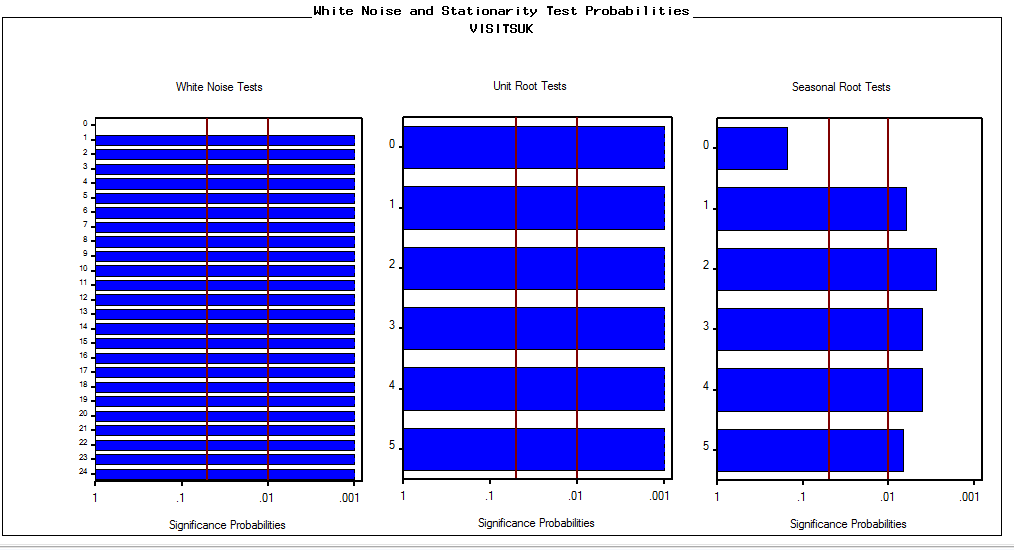
## Data Exploration:

We performed preliminary data exploration to identify trends and seasonality in the data.

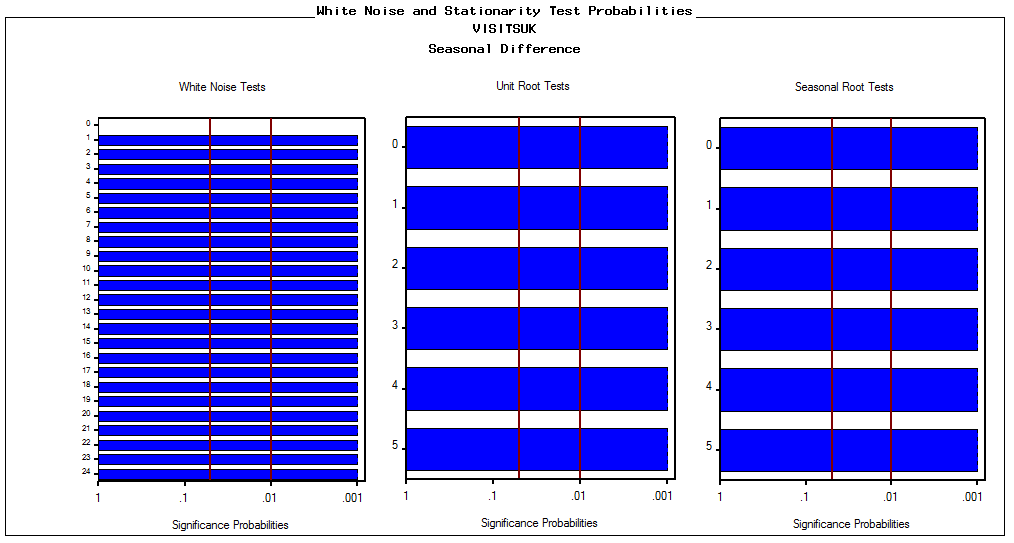


The above graph for the target variable ***VISITSUK*** indicates the presence of seasonality in the monthly data. And overall there seems to be a positive linear trend over the time. The auto correlation plots and test plots reveal the following

1. ACF resembles a sinusoid with decaying amplitude and the plots were significant at lag 12.
2. There is Stationarity in the series
3. Seasonality results shows that there is only partial seasonality available.  
   



Since we had already observed seasonality pattern in the series we apply seasonal differencing. After seasonal differencing the original data, we see good results in seasonality root tests indicating presence of seasonality pattern in the data.



Similarly, for the target variables ***Visits to Abroad, Earnings in UK and Expenditure Abroad*** we performed similar analysis to build independent candidate models. Please refer appendix section for detailed explanation.

## Approach:

We used standard SEMMA approach to build our forecast models. Since this is a time series data, there is no necessity of Sampling the data. Data is already a measured statistic of the population.

**Training, Validation and Forecasting horizons:**

Initially we divided the monthly time series data for the variable ‘***visitsuk’*** into training and validation as follows: The hold out sample was considered as there was no exceptional behavior within the sample at the end of the data series.

* The fit sample is taken from January, 1980 to October, 2010
* The hold out sample range November, 2010 to October, 2015

The forecast horizon is taken from November, 2016 to October, 2018.

## Modelling:

To extrapolate past behavior into future, we used random walk models, auto regressive models, moving average models, trend models, seasonal models, smoothening exponential and ARIMA time series models.

We have selected MAPE as the accuracy statistic to our business problem.

As a part of descriptive and exploratory analysis we need to identify the trend and seasonality in the data. To diagnose the trend and seasonality we looked at the interactive data graphs, ACF, PACF and IACF plots.

Following are the points we considered while building forecasts:

* For the stationary time series with trend, if there are autocorrelations in ACF, PACF and IACF plots then, we applied basic ARMA models along with trend curve models
* For a non-stationary time series with trend, we performed simple differencing and then added ARIMA stationary models to pass white noise and unit root tests
* For a stationary time series with seasonality, if there are autocorrelations we applied basic ARMA models along with deterministic seasonal components
* For a non-stationary time series with seasonality, we performed differencing and then added ARMA stationary models to pass white noise and unit root tests
* For some time series data, to model trend and seasonality we combined the respective components with ARMA stationary component and modelled forecasts
* Once we built the models, we filtered the top models based on the accuracy statistic, residual errors, parameter estimates and confidence intervals of forecasted values for further analysis
* Considering the best performing model from the previous step, we used the cross correlation function to check for any correlations of target variables with all the explanatory variables by checking the significant lags in the cross correlation graphs. If there is any significant impact from the graph, then we decided to add that explanatory variables as the dynamic repressors to the existing best fit model in order to improve accuracy.
* For time series which had significant values at specific lags, we applied factored ARIMA model (also referred to as a multiplicative model) as a product of simpler ARIMA models for better treatment of residuals
* If there are any sharp spike or dips in the time series data for the target variable, we checked for interventions by doing basic research and then introduced interventions if needed
* We also combined the two best models using combine models option in SAS forecasting module, to see if there is any improvement in the performance of the model. This is also called as Ensemble modeling.
* Once we fit the sample data and decide on the final forecasting model, then we evaluated the model accuracy using the hold out sample

Cross Correlation Function (CCF):

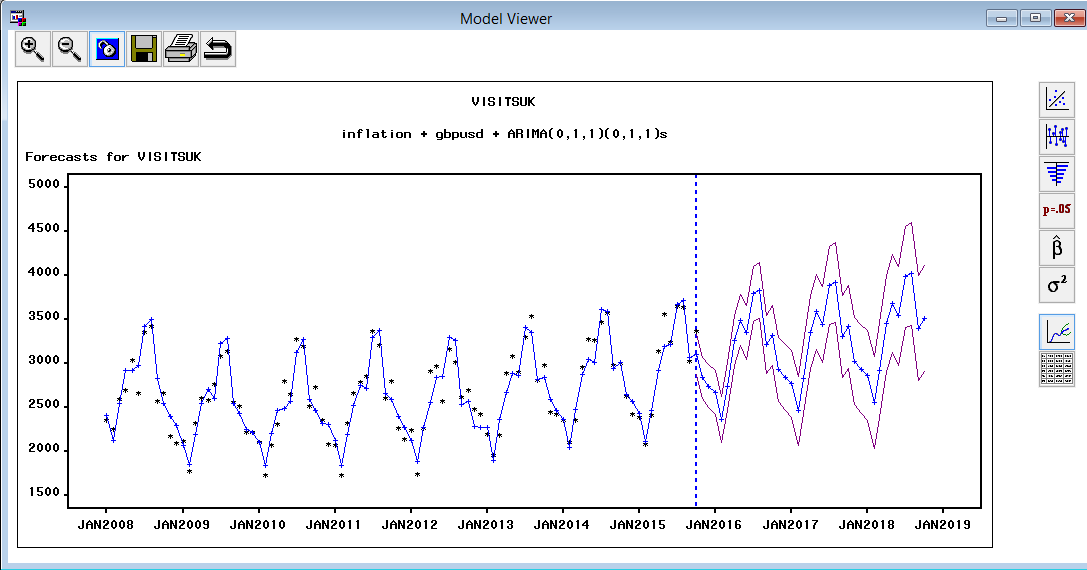
Characterizing a time series is not just estimating mean and standard deviation but we had also considered the cross correlation function to model a series in order to include the effects of past and current values of other series in the model. The ‘**CROSSCORR’** function in SAS computes the cross-correlations of the changes in Y with the changes in X.

The SAS code for cross correlation function which we used in building candidate models is embedded in the appendix section as ‘crosscorrelation.sas’

## Forecasting

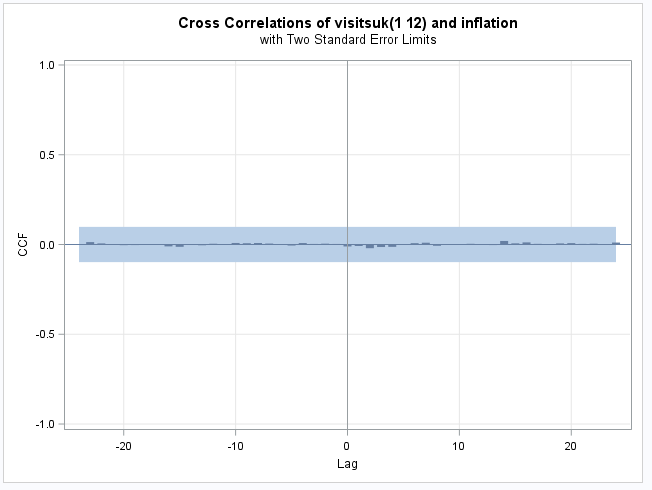
Using the above methodology, we did modelling for UK inbound and outbound – Visits and Spending’s

For ***Monthly Inbound Visits to UK***, following is identified as the best model

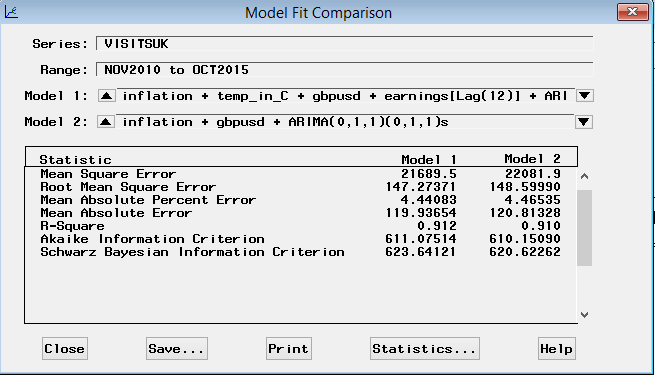


We saw that the above model is good with residual analysis and it passes both the white noise and stationarity tests. It is also observed that the parameter estimates are not significantly different from zero and with MAPE of 4.46% (Please refer appendix for this model characteristics)

We had also checked for the cross correlations of ***visitsuk*** variable with other input series variables. The cross-correlation graph for visitsuk with respect to inflation is shown below. It can be inferred from the graph that there were no significant lags and therefore we haven’t included it as dynamic repressor. (Please refer to appendix for the other CCF graphs for visitsuk variable with respect to other input series variables (usdgbp, oilgbp, temp\_in\_C, gbpusd)



Comparison across 2 best models for ***visitsuk***

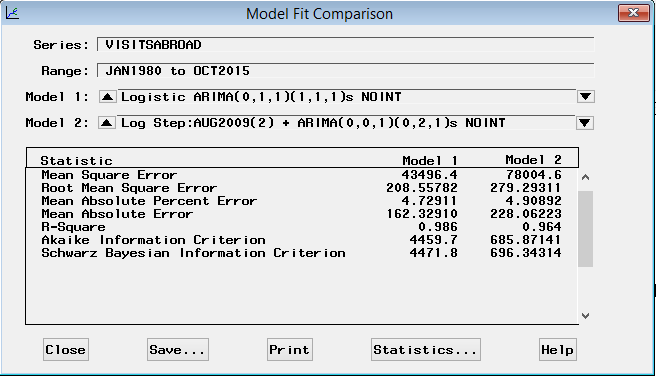


For ***Monthly Outbound Visits from UK***, following is identified as the best model

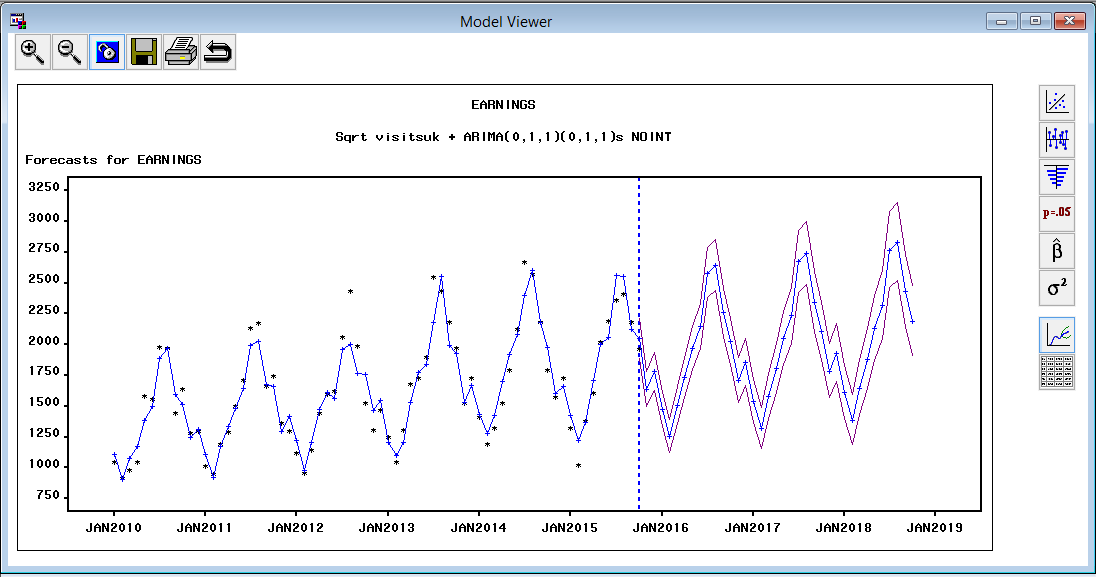


We had seen that the above model is good with residual analysis and it passes both the white noise and stationarity tests. It is also observed that the parameter estimates are not significantly different from zero and with MAPE of 4.72%. (Please refer appendix for this model characteristics)

Comparison across 2 best models for ***visitsabroad***

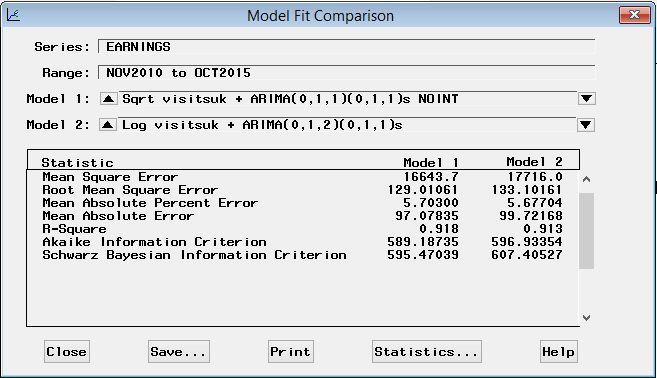


For ***Monthly Inbound Earnings to UK***, following is identified as the best model

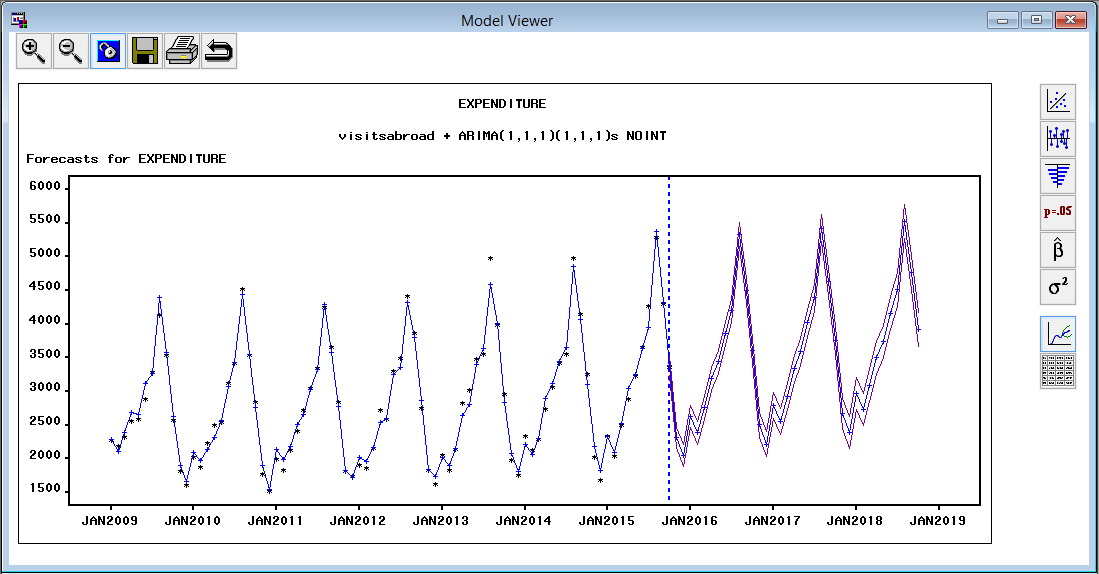


We saw that the above model is good with residual analysis and it passes both the white noise and stationarity tests. It is also observed that the parameter estimates are not significantly different from zero and with MAPE of 5.70%. (Please refer appendix for this model characteristics)

Comparison across 2 best models for ***Earnings***

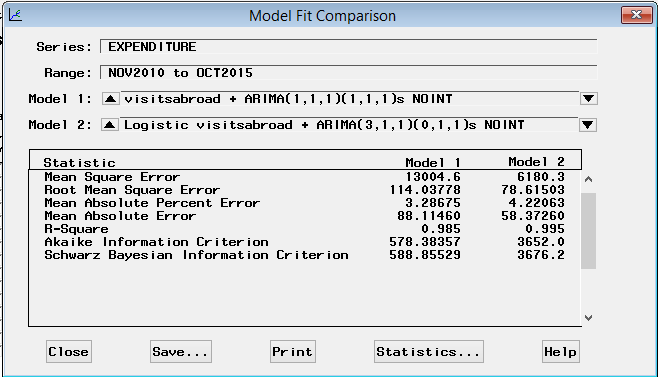


For ***Monthly Outbound Expenditure from UK***, following is identified as the best model



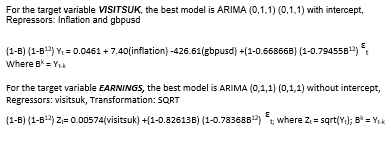
We saw that the above model is good with residual analysis and it passes both the white noise and stationarity tests. It is also observed that the parameter estimates are not significantly different from zero and with MAPE of 3.28%. (Please refer appendix for this model characteristics)

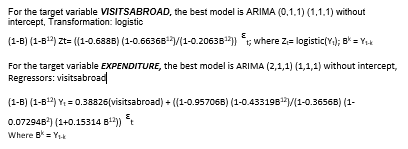
Comparison across 2 best models for ***Expenditure***



## Model Equations:

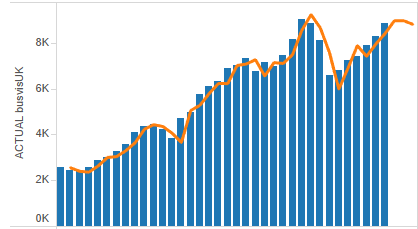
The forecast equations for the above 4 models are:



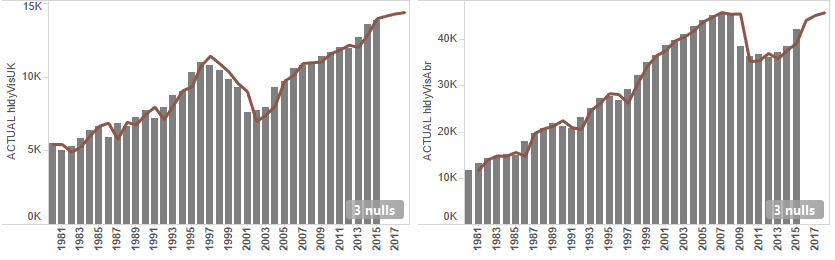


# Important Observations from Forecasts:

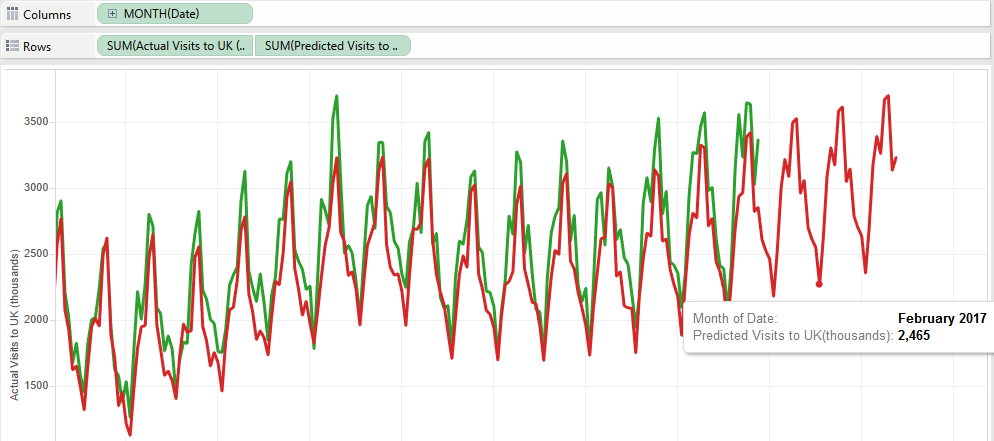
* In the below forecasted graph of yealry business visits to UK, we observe a 2% decrease in number of business visits to UK during 2017.



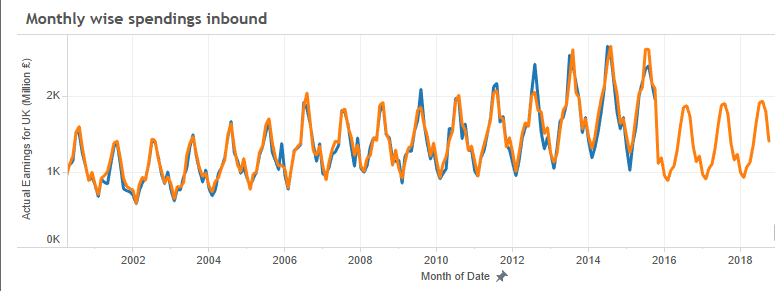
* In the below forecasts of Inbound and Outbound holiday tourism graphs, we observe that number of outbound visitors from UK are almost 3 times more than that of inbound visitors who travel to enjoy their holiday.



* In the below forecast of inbound monthly visits to UK, we observe that there is a trend and seasonality continuously increasing as time flies by. We expect the number of tourists visiting UK will be less during and after Christmas season and will be high during and after Summer.



* In the below forecast graph of monthly spending in UK by visitors/tourists, we observe that the inbound spending from tourists is reduced by almost 25% over the forecasted time.

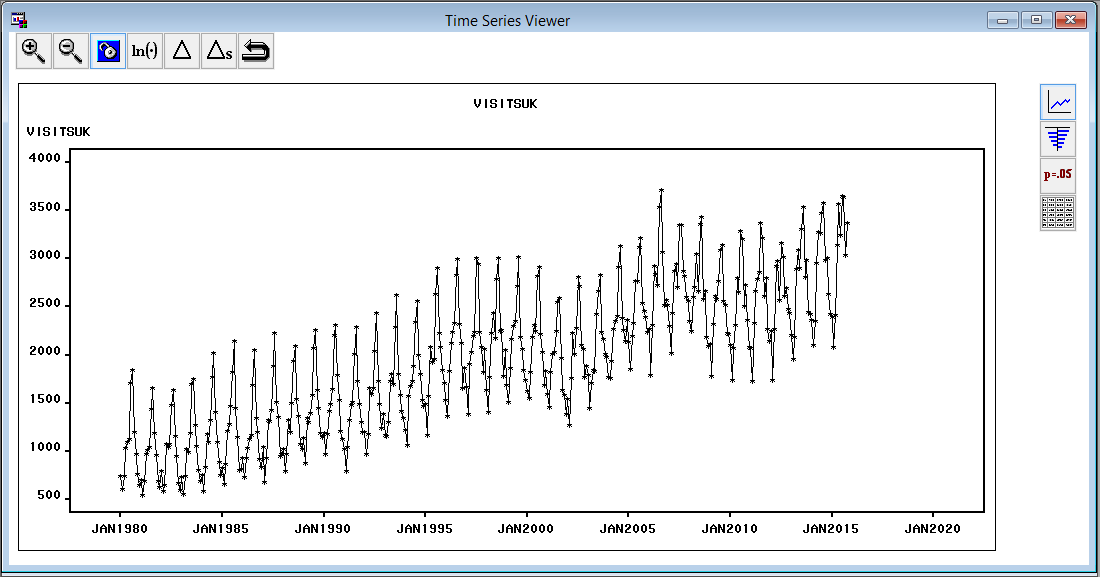


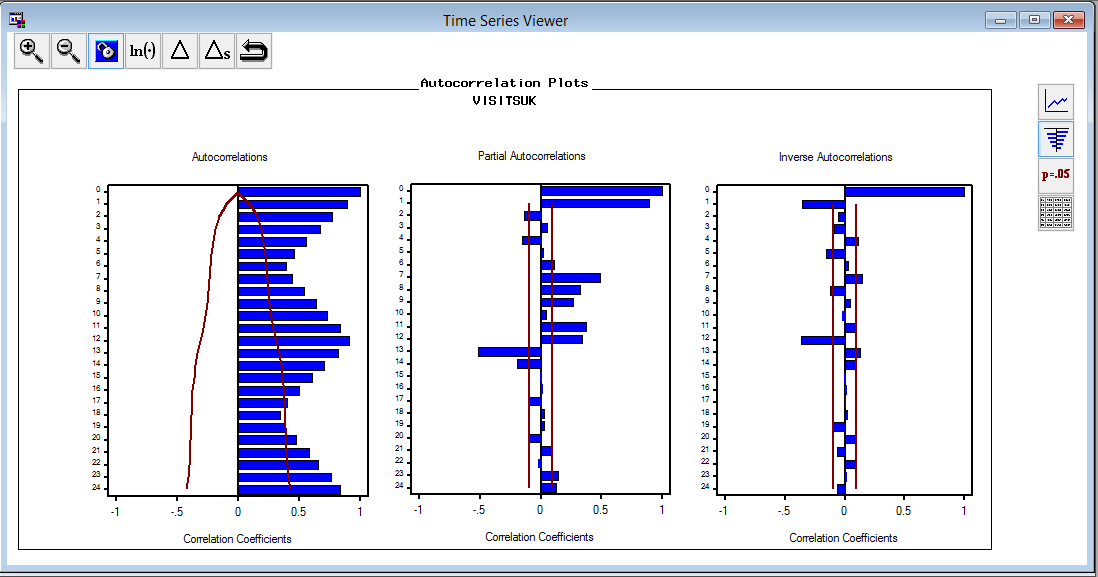
# Conclusion/Recommendation:

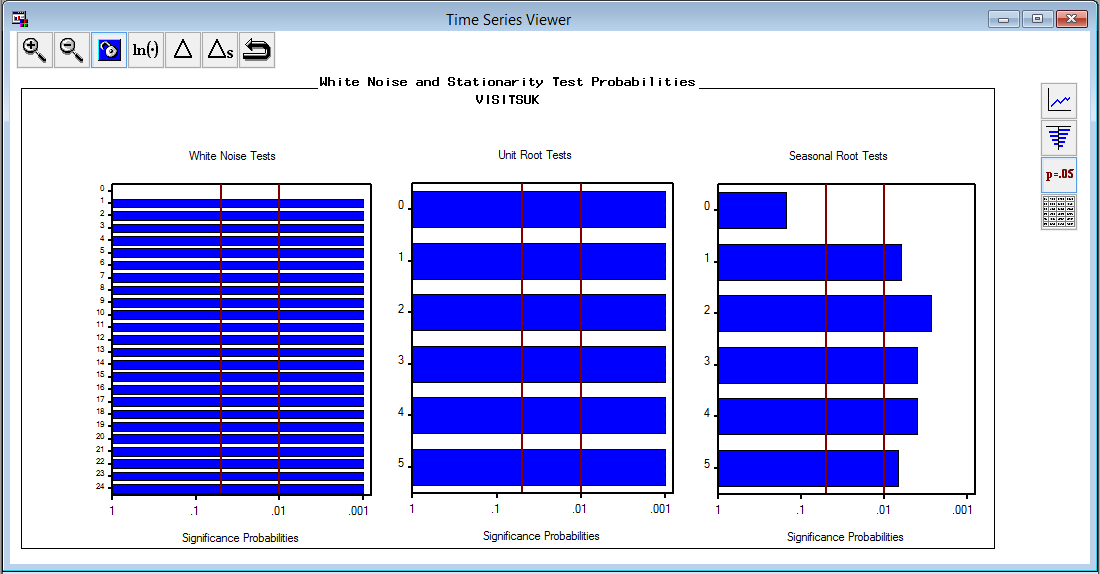
* From our monthly inbound spending forecasts, we find that there is a decreasing trend over the next 3 years. We expect this drop to be as severe as 25% with Mean Absolute Percent Error of 4.46% in the forecasted model. So we recommend UK tourism department to bring in more events/attractions to lure/sustain the tourist population.
* There is a seasonal trend of low number of visits during and after Christmas season (December to February) so we suggest to offer seasonal discounts on tourism packages to attract more tourists and change the past seasonal trends.
* On yearly purpose of visits data, we forecasted that inbound tourism on business purposes might drop by 2% in 2017. So we suggest UK government to conduct more business summits or events in UK to draw business visitors from all parts of the world. Business visits tend to generate more revenue than any other reason.
* Also on yearly purpose of visits data, we forecasted that inbound visits on a holiday will be very less compared to holiday outbound visits (1:3). We recommend UK tourism department to cash this opportunity by designing new international holiday packages, so that these outbound expenditures can be spent effectively and efficiently by the government itself.

# Appendix

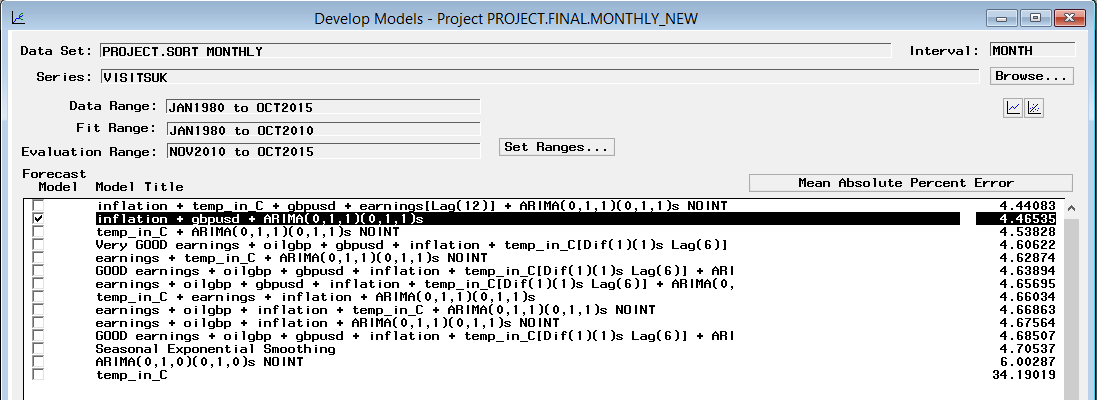
**Initial Data Exploration of Monthly inbound visits to UK:**



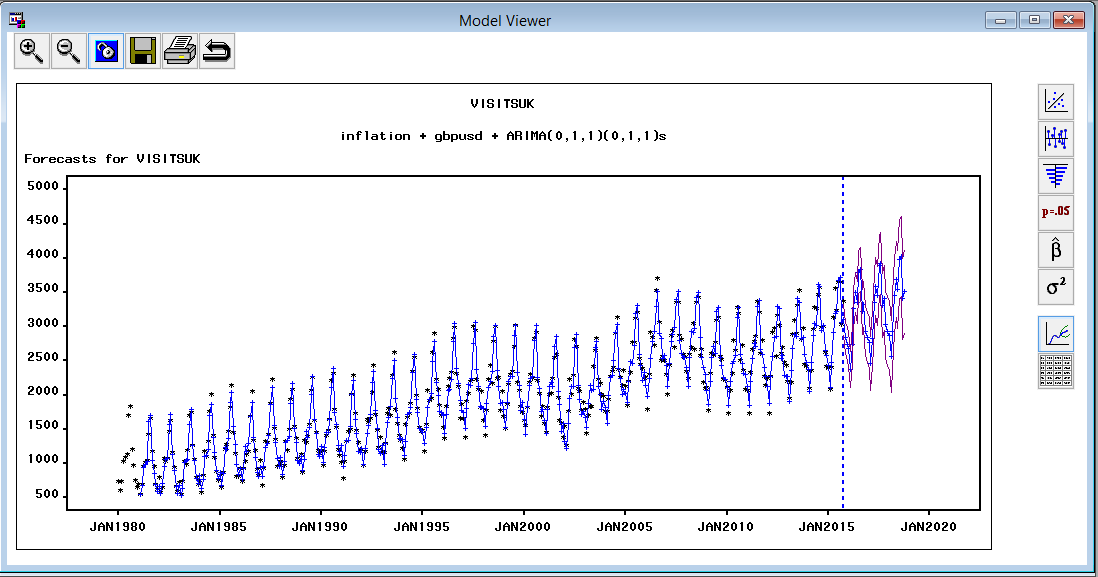


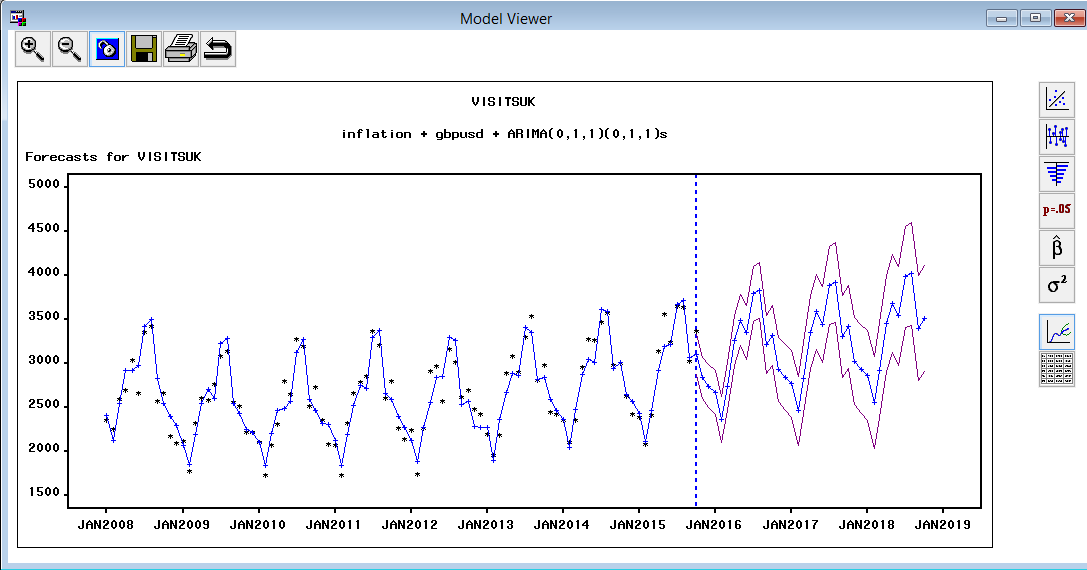


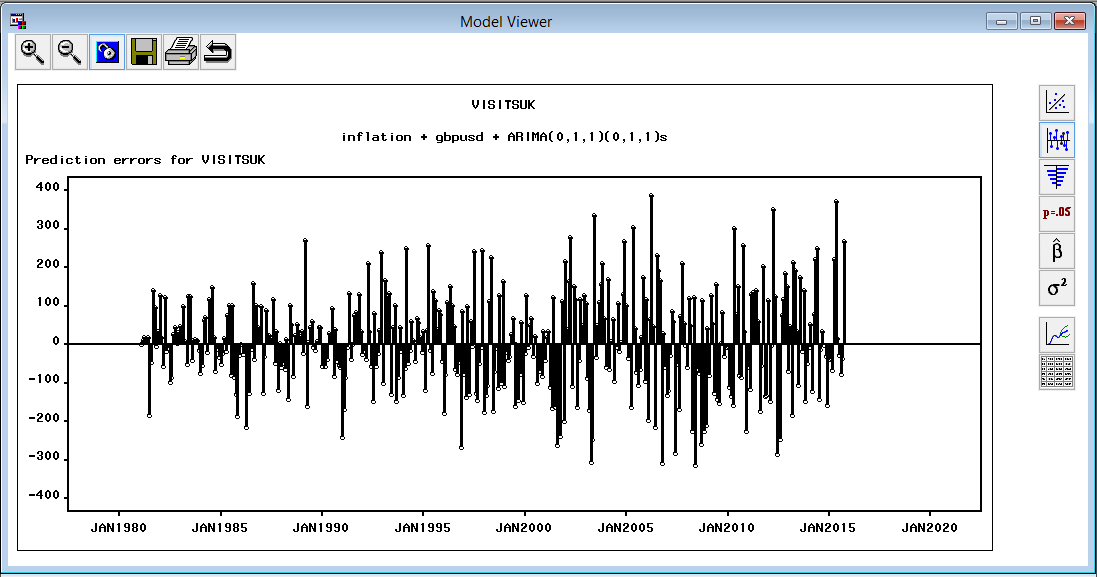
**Candidate models for Monthly inbound visits to UK:**

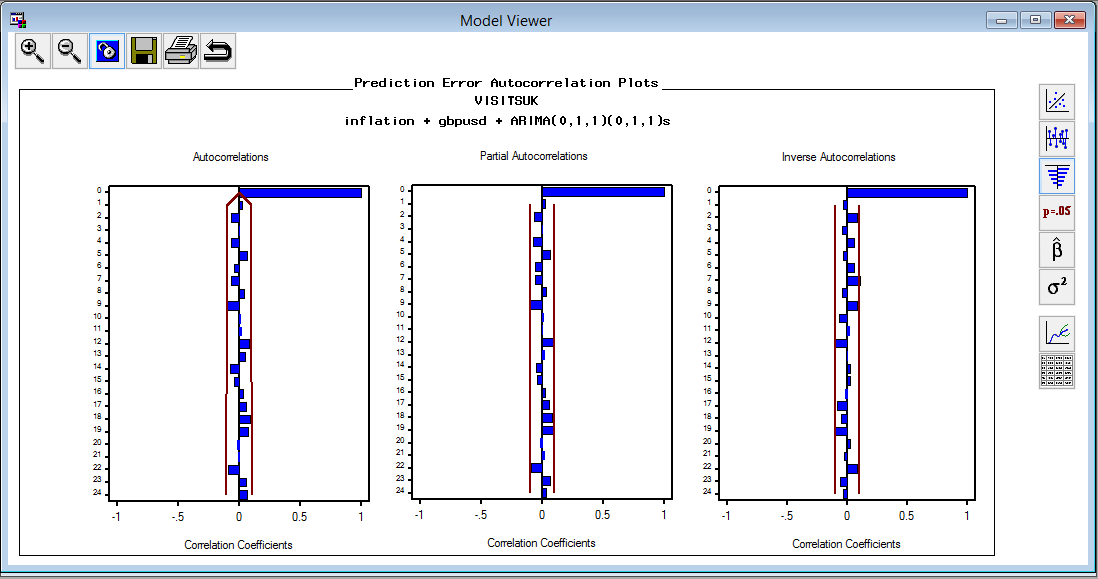


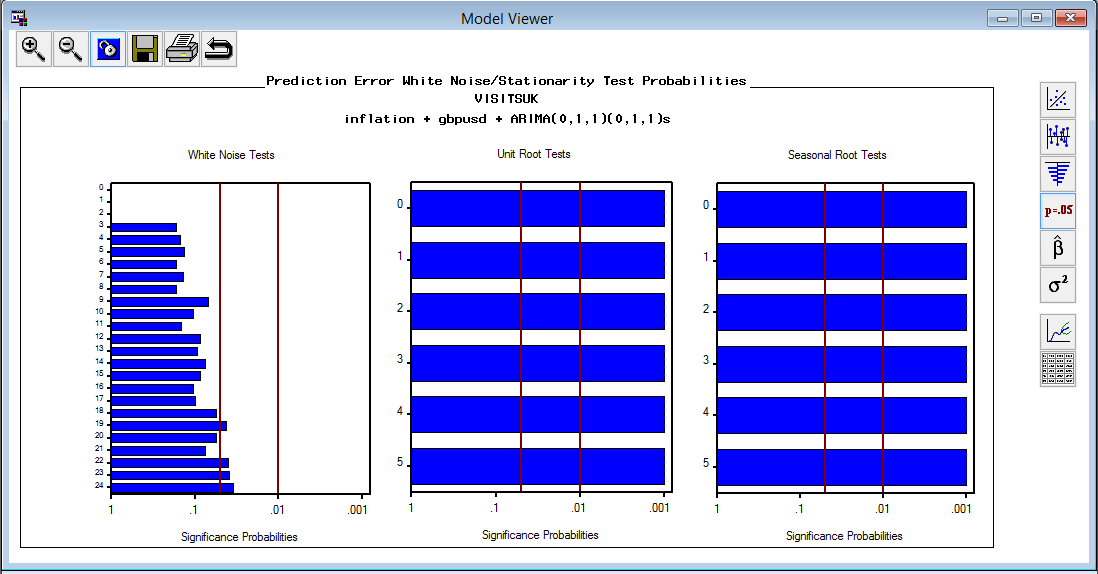
**Forecasted Monthly Inbound Visits to UK**

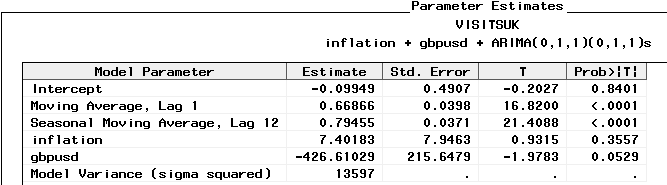


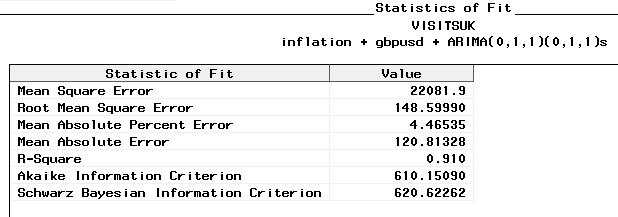




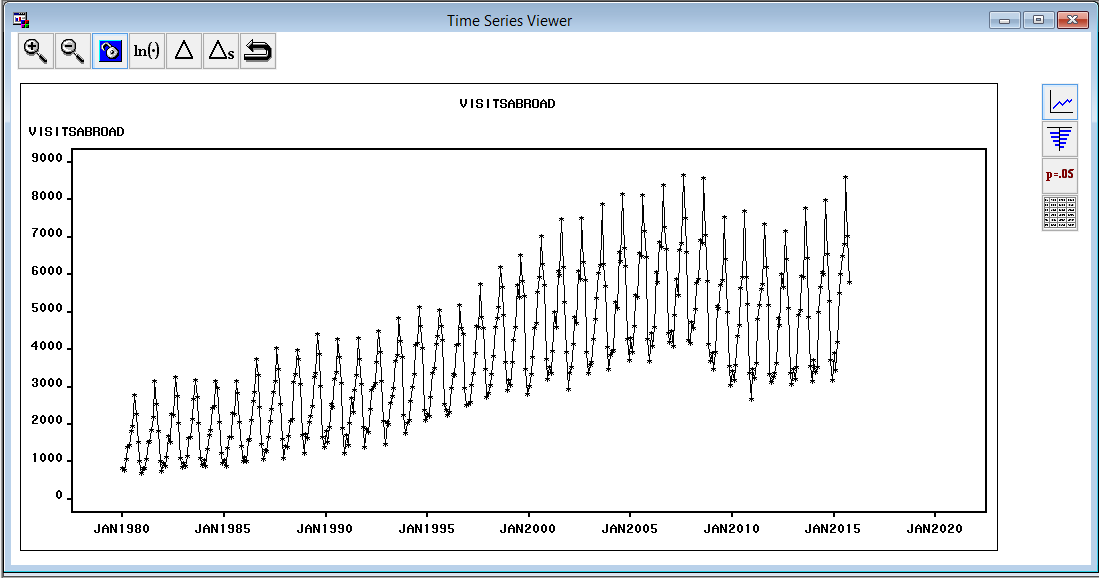


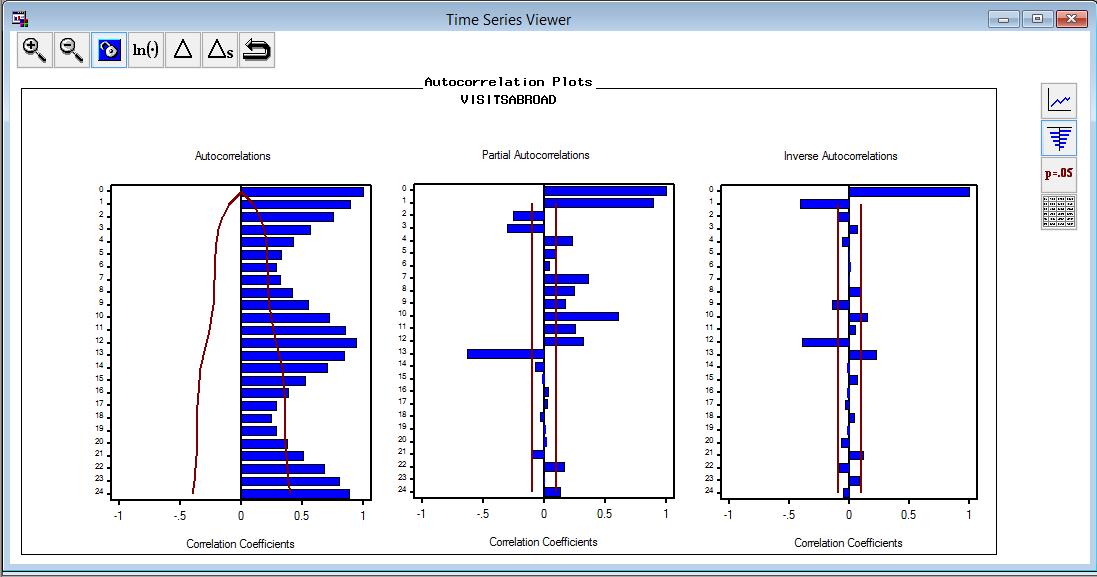


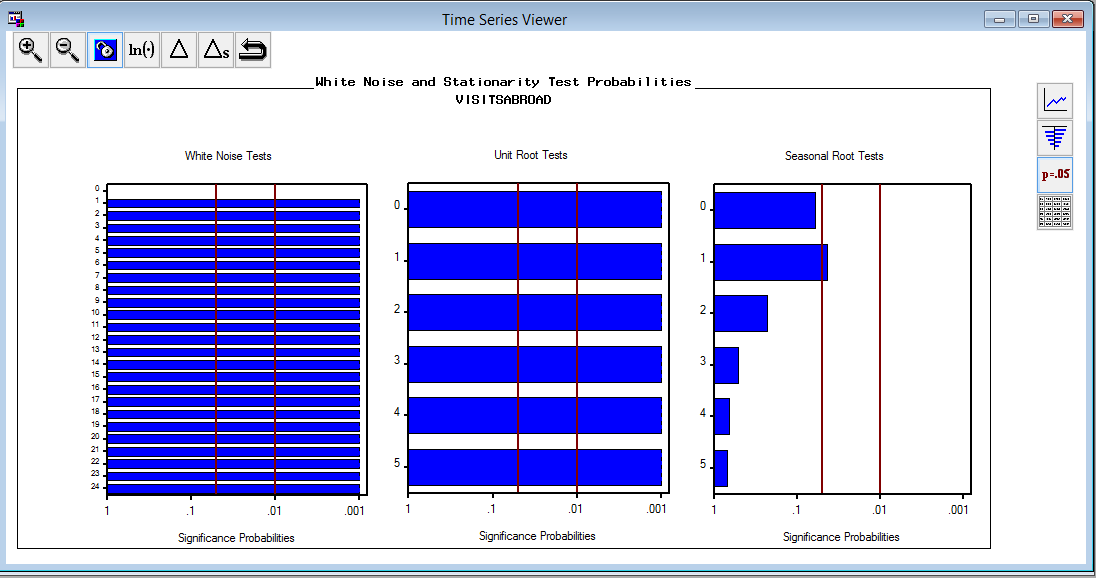


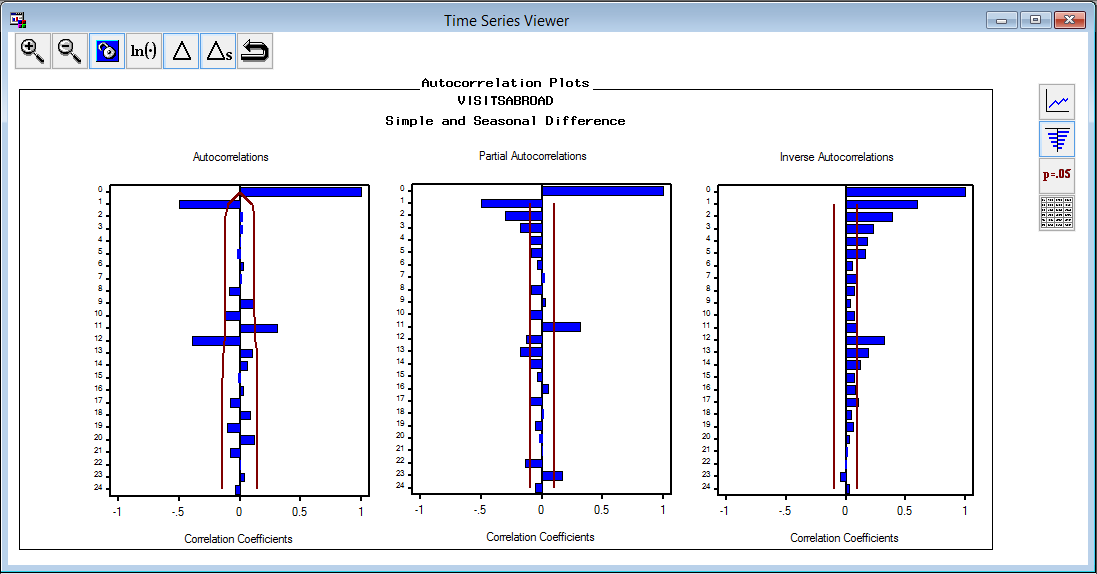


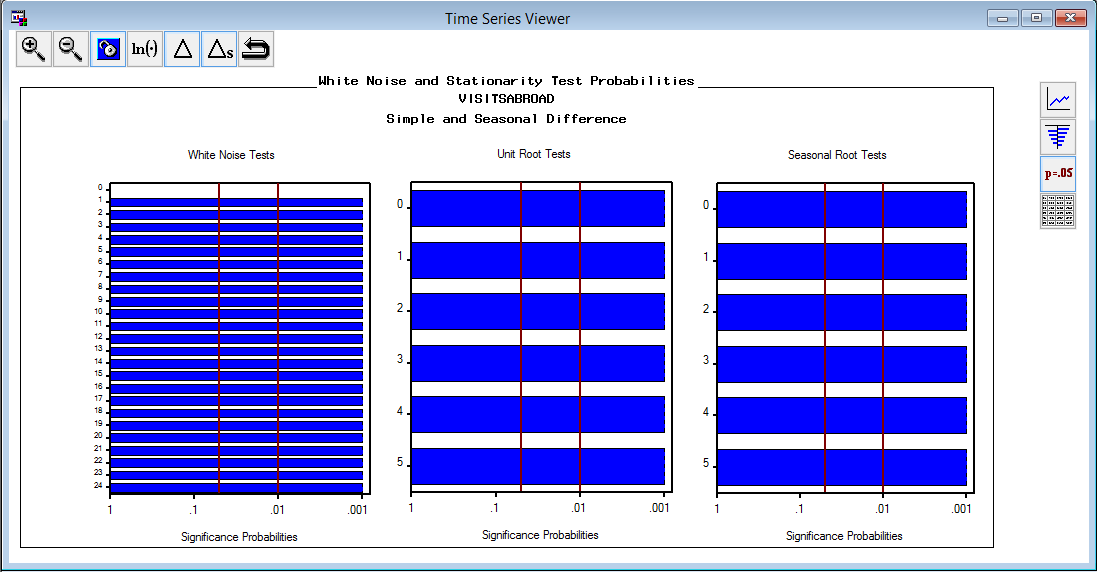
**Initial Data Exploration of Monthly Outbound visits from UK to Abroad:**



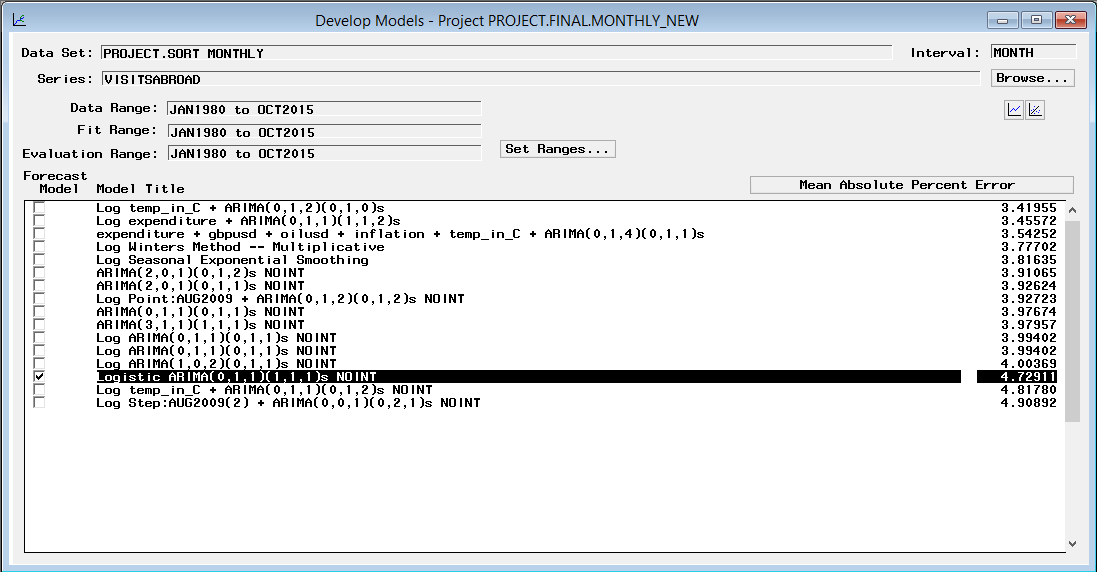




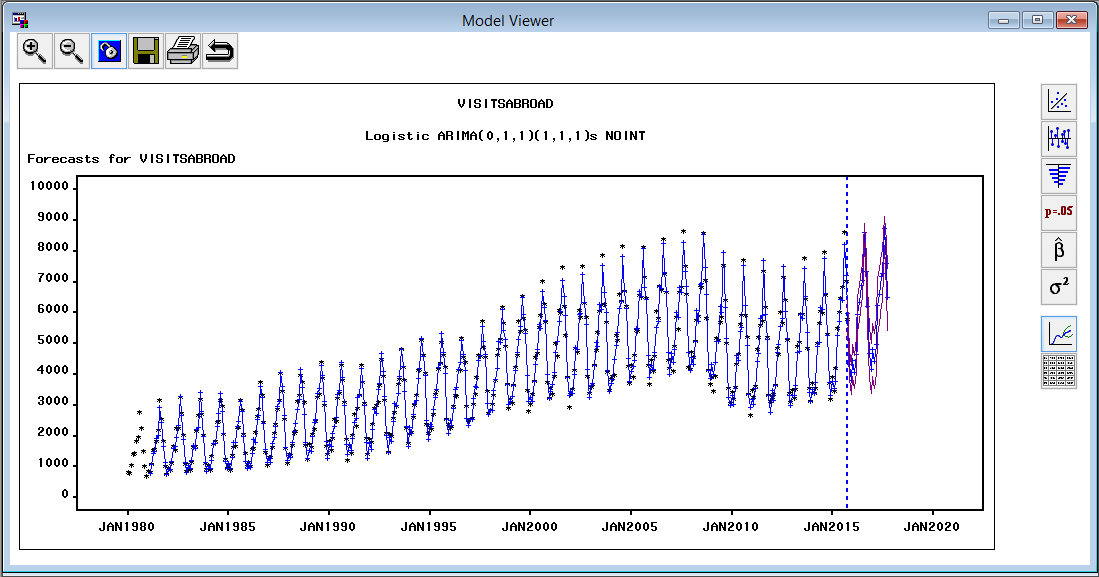




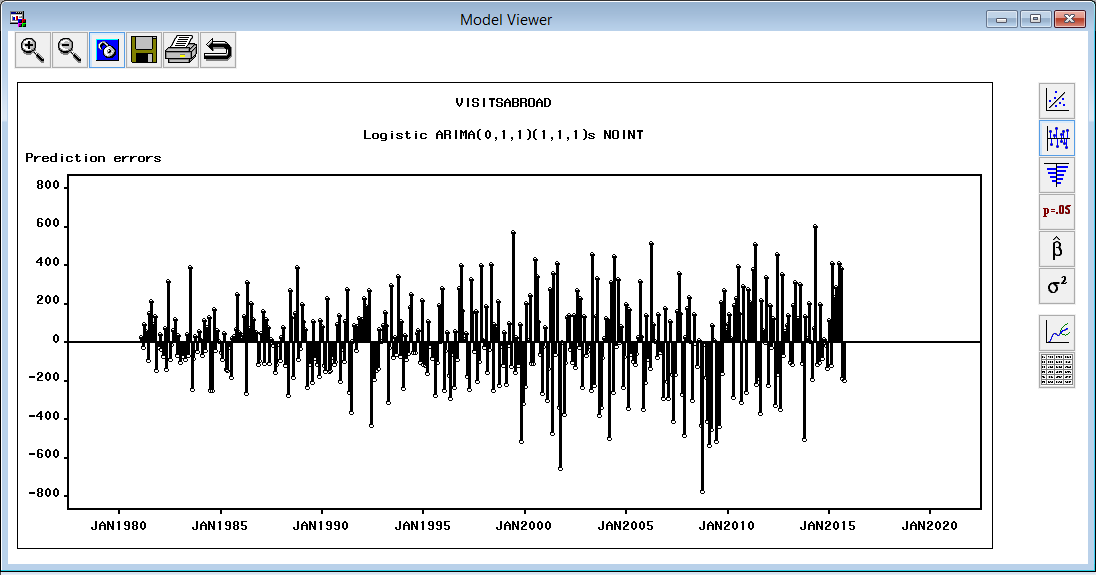
**Candidate Models for Monthly Outbound visits:**

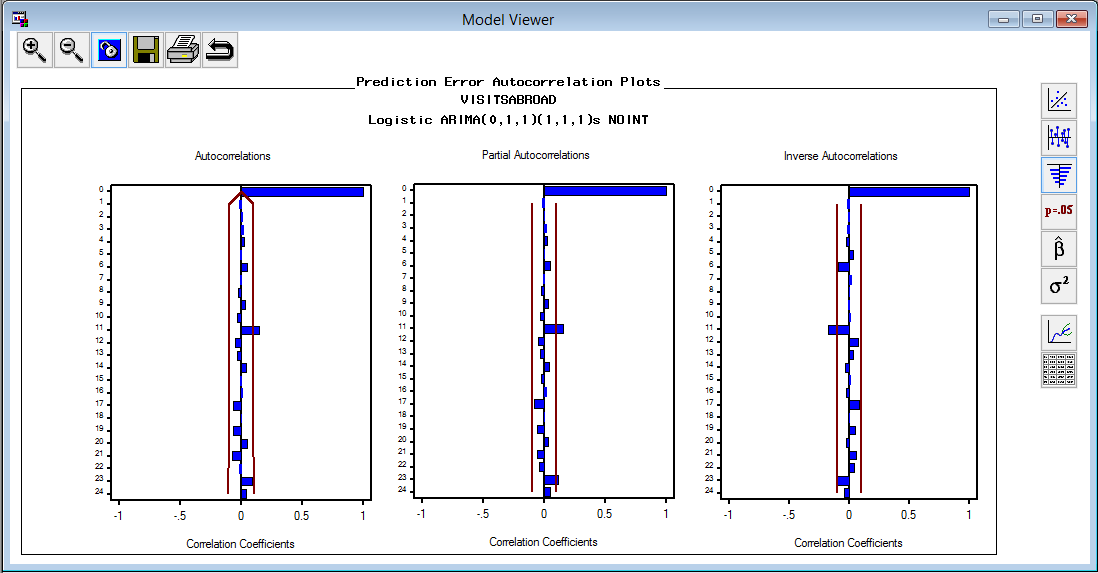


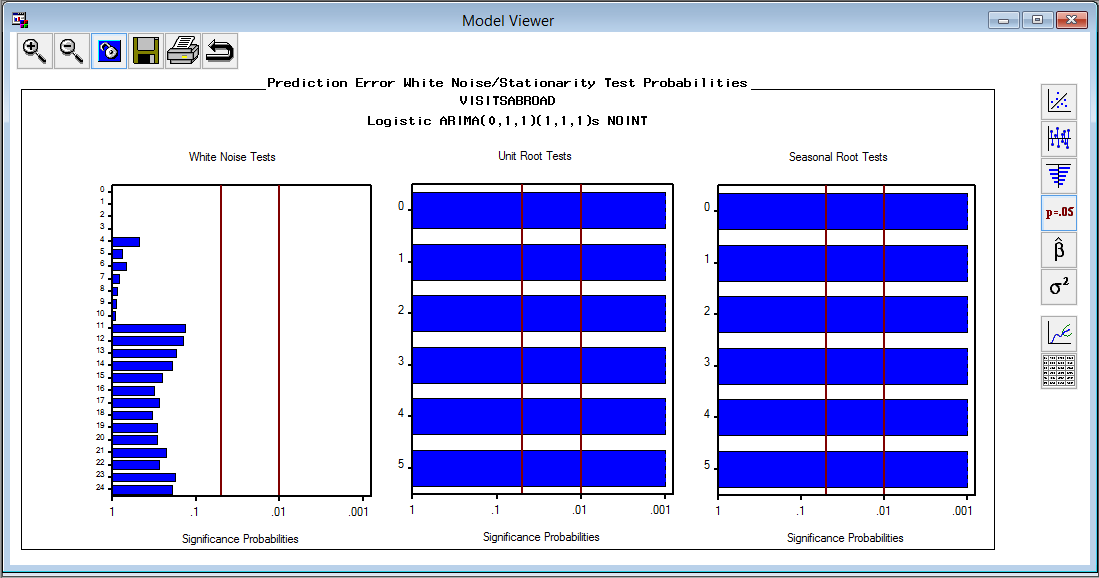
**Forecasted Monthly Outbound Visits to Abroad:**

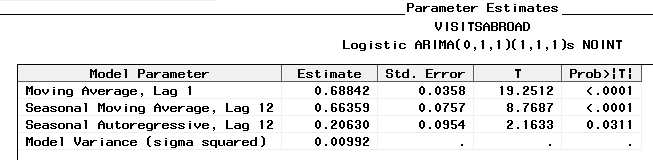


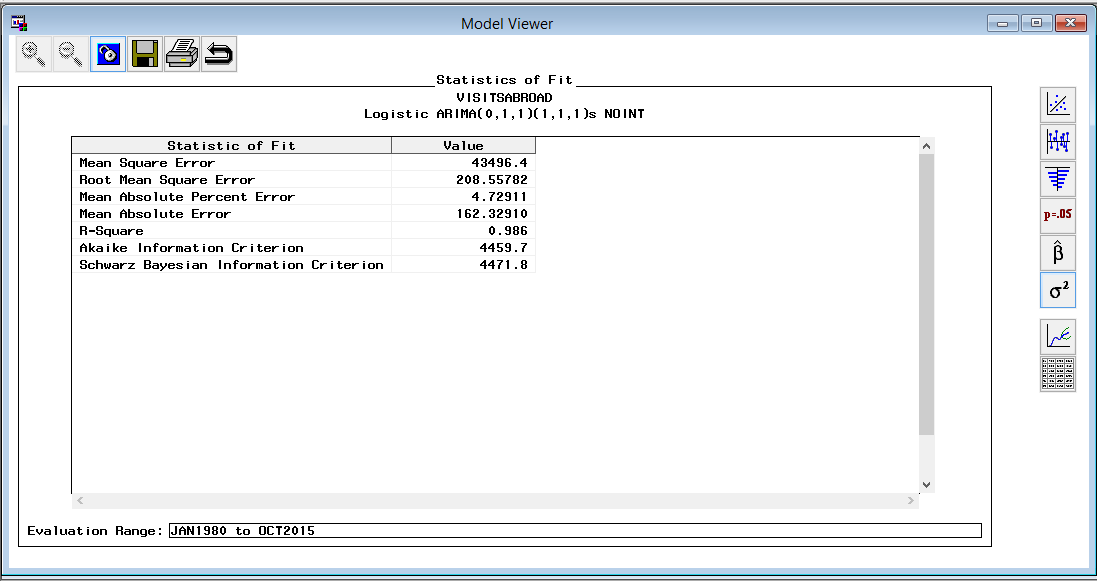




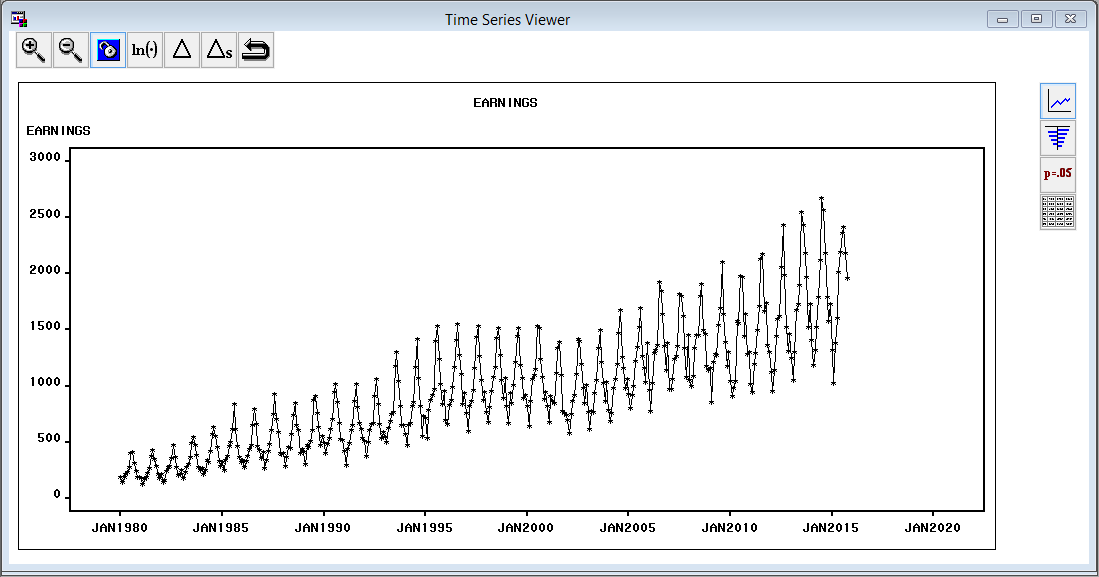


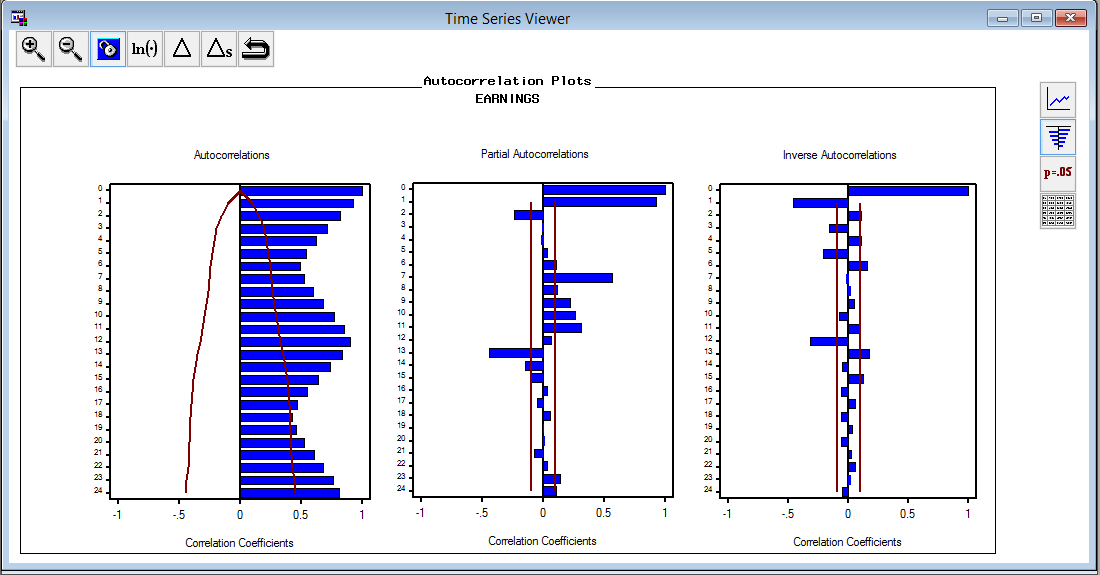


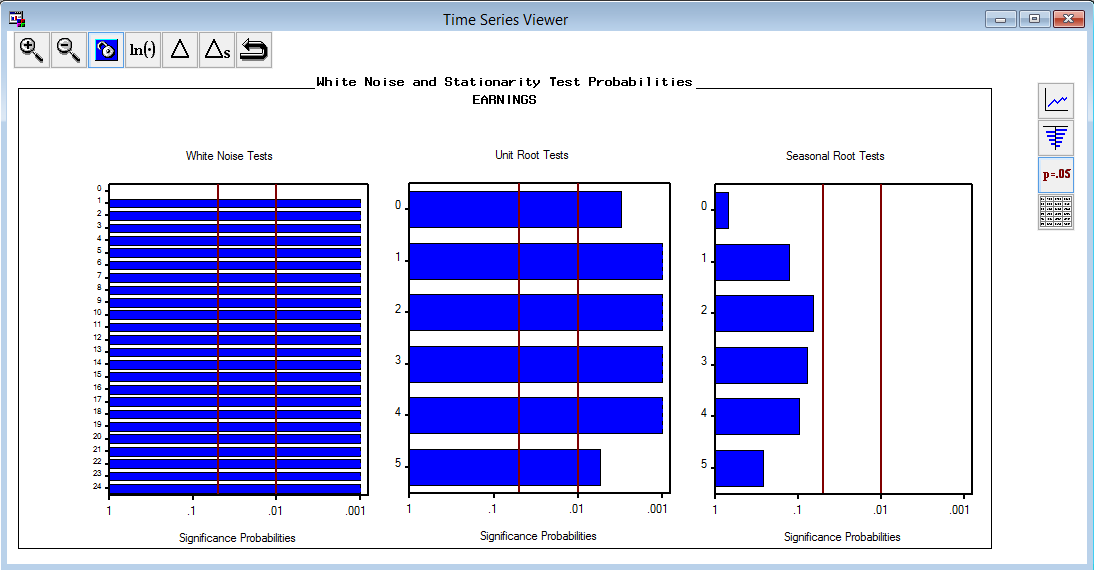


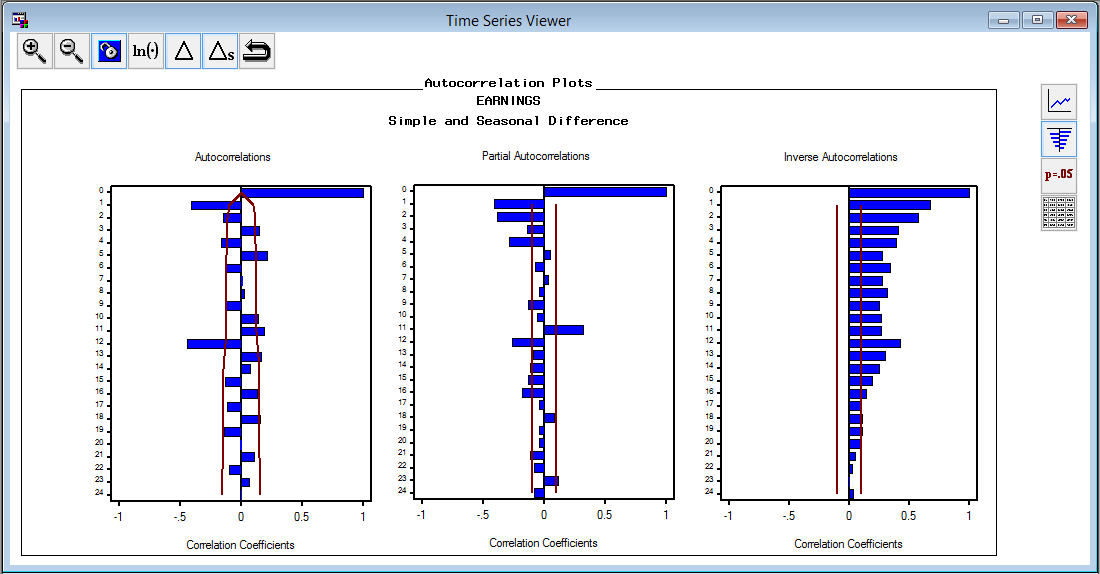


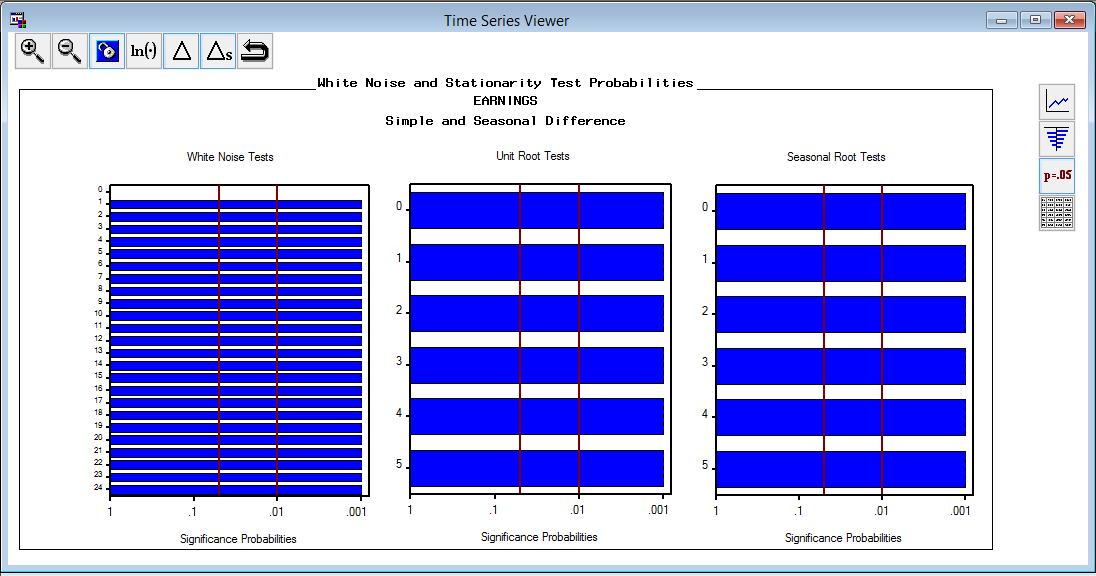
**Initial Data Exploration of Monthly Inbound Earnings to UK:**



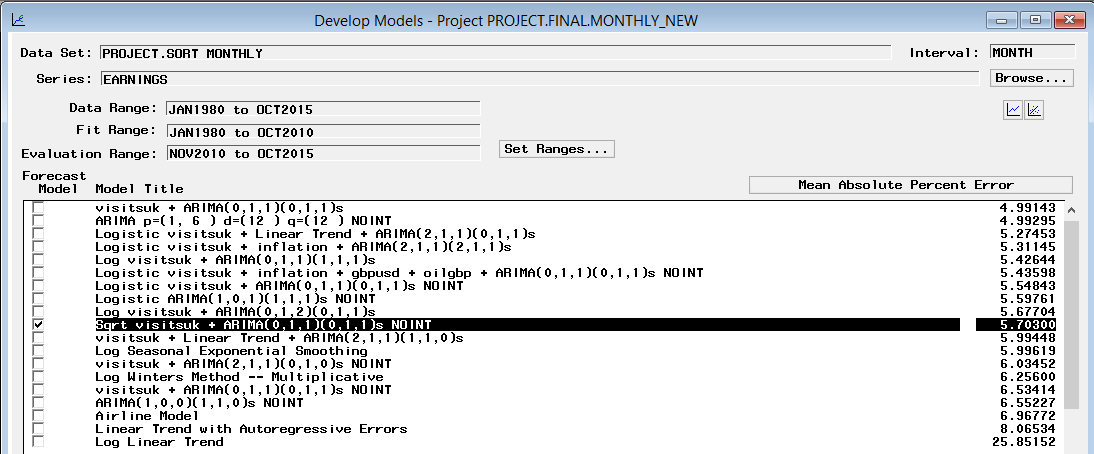




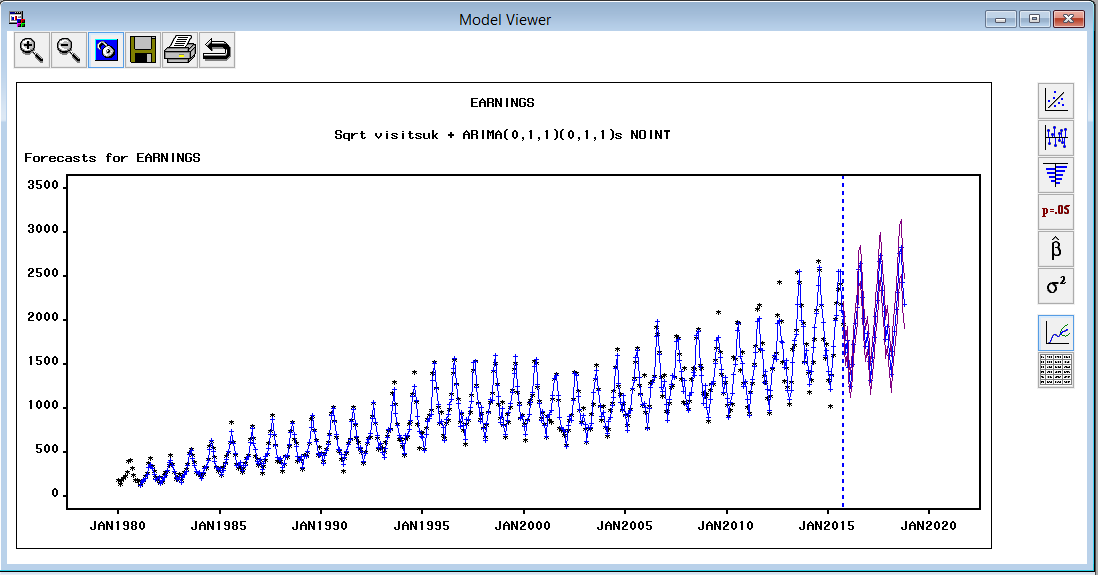


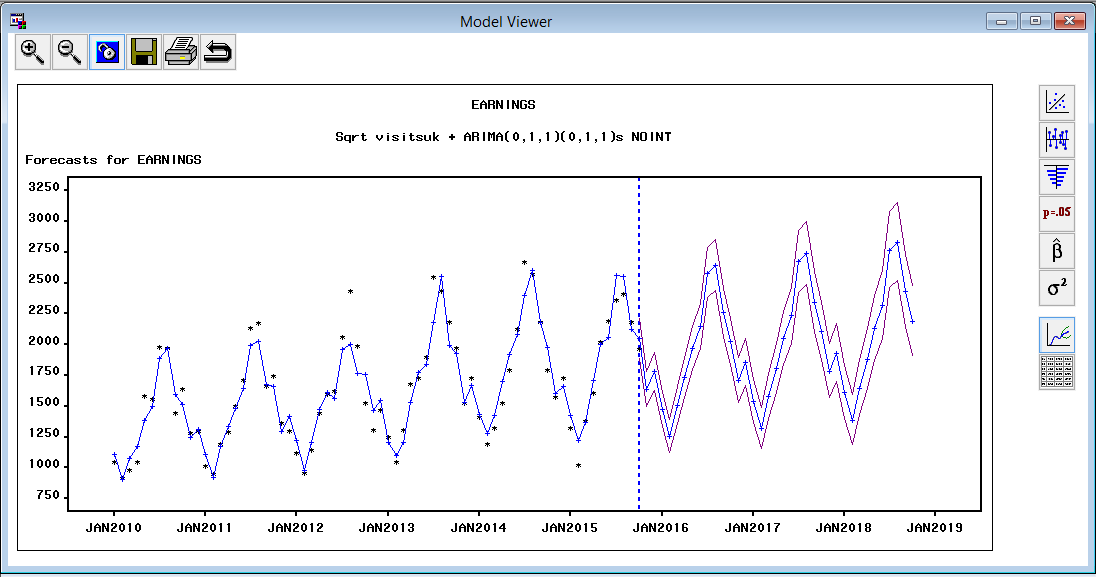


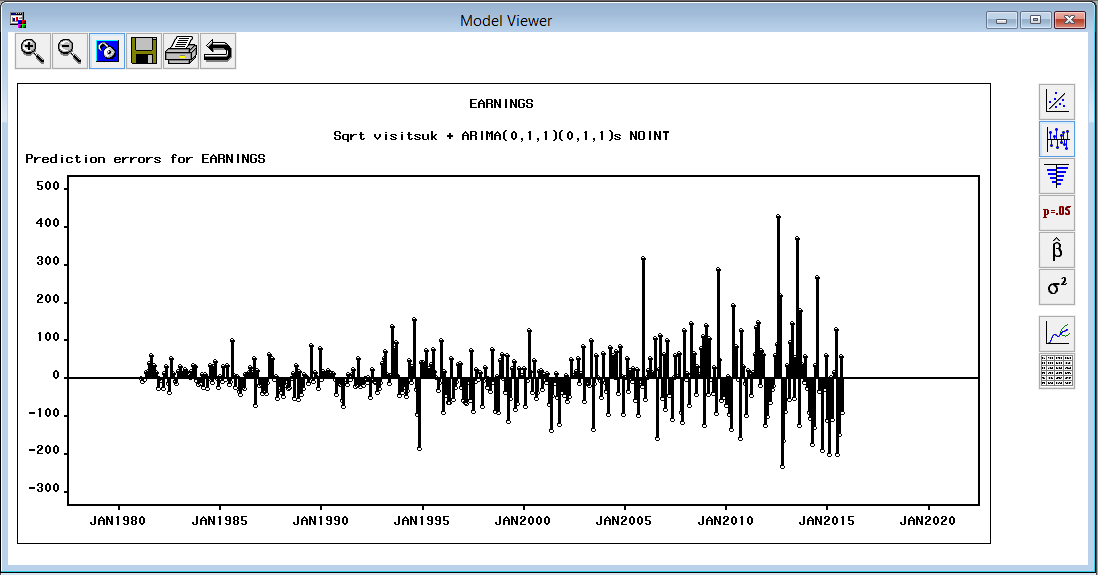
**Candidate models for Monthly inbound earnings to UK:**

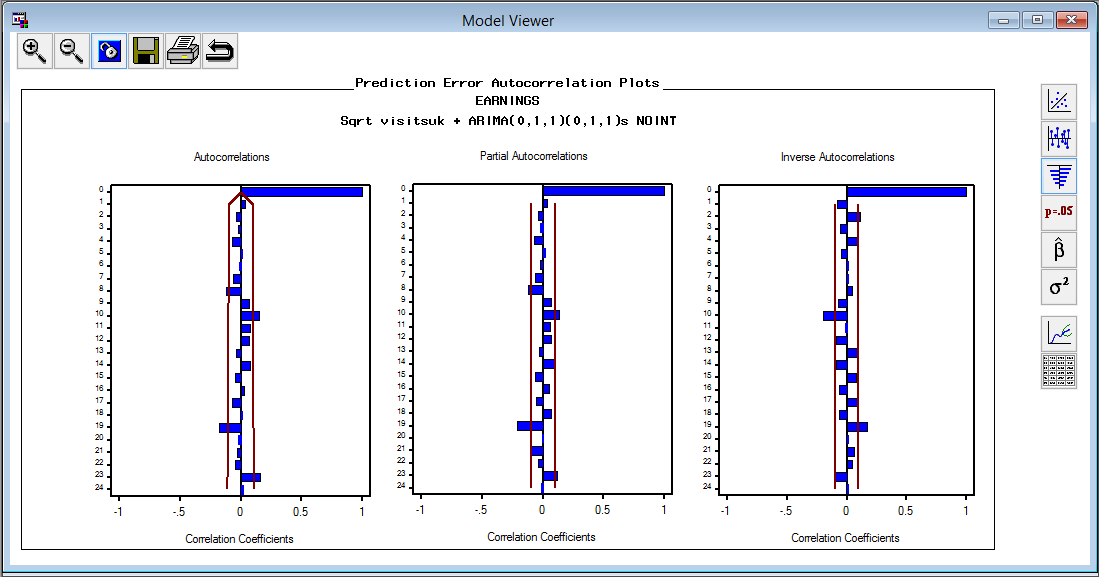


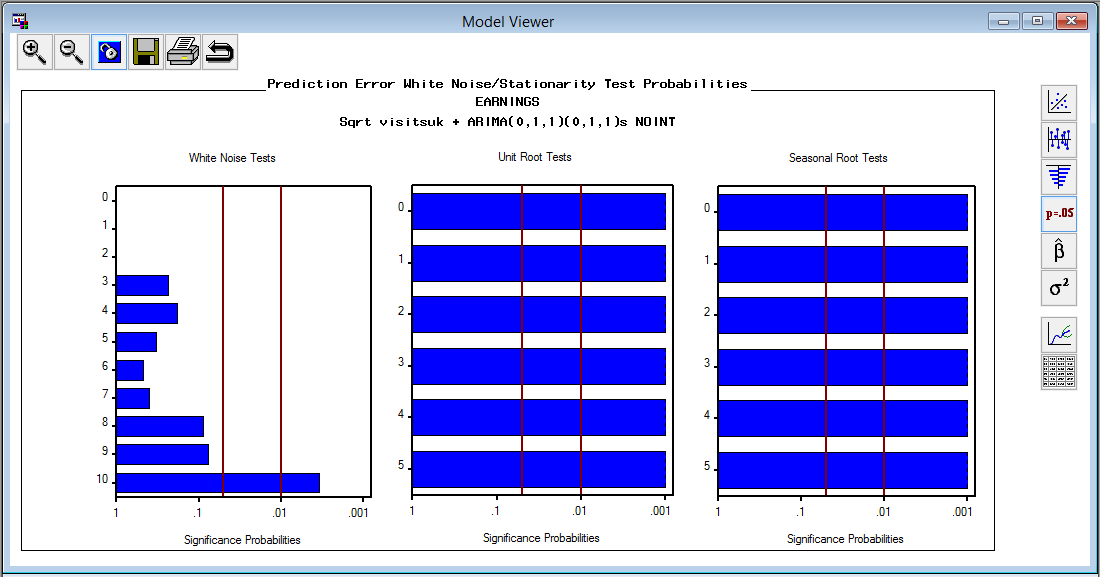
**Forecasted Monthly inbound earnings to UK:**

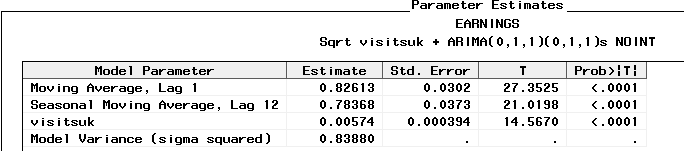


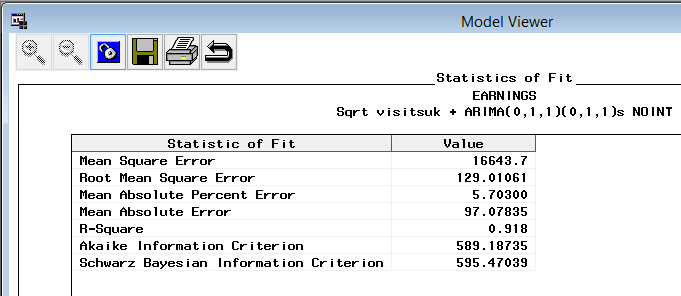




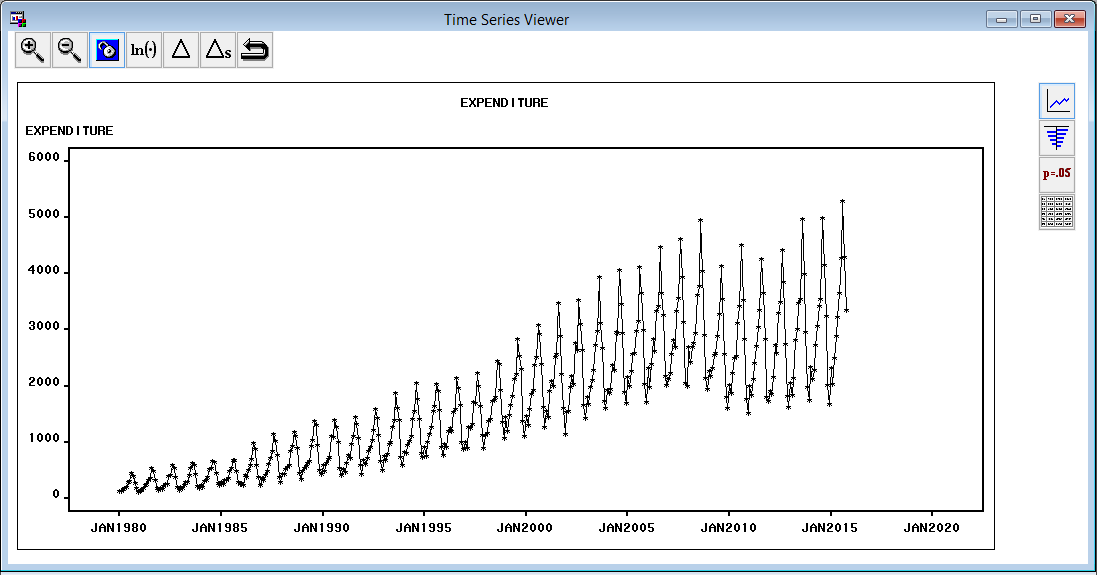


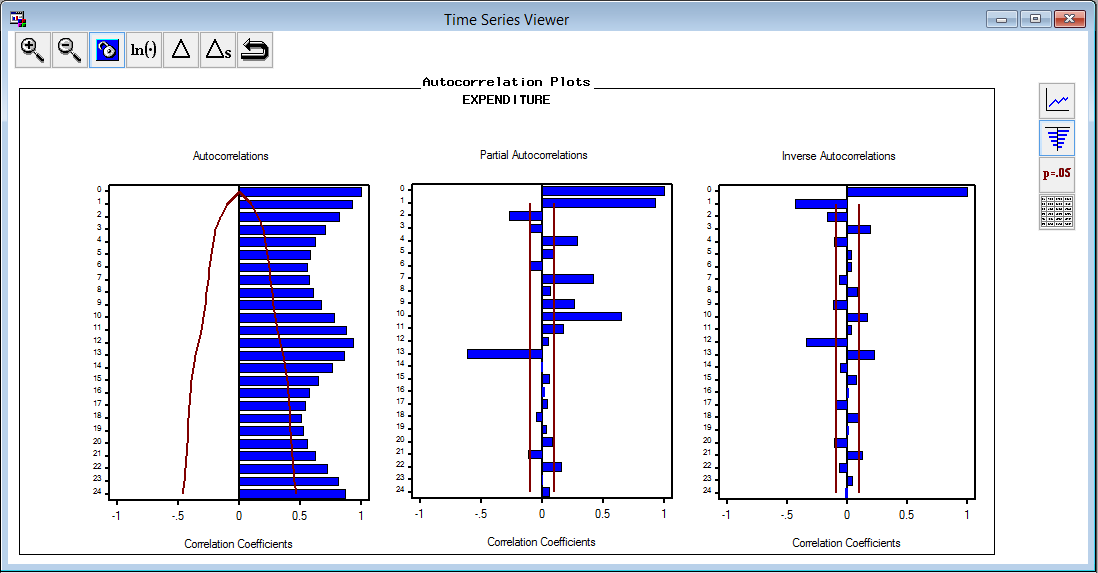


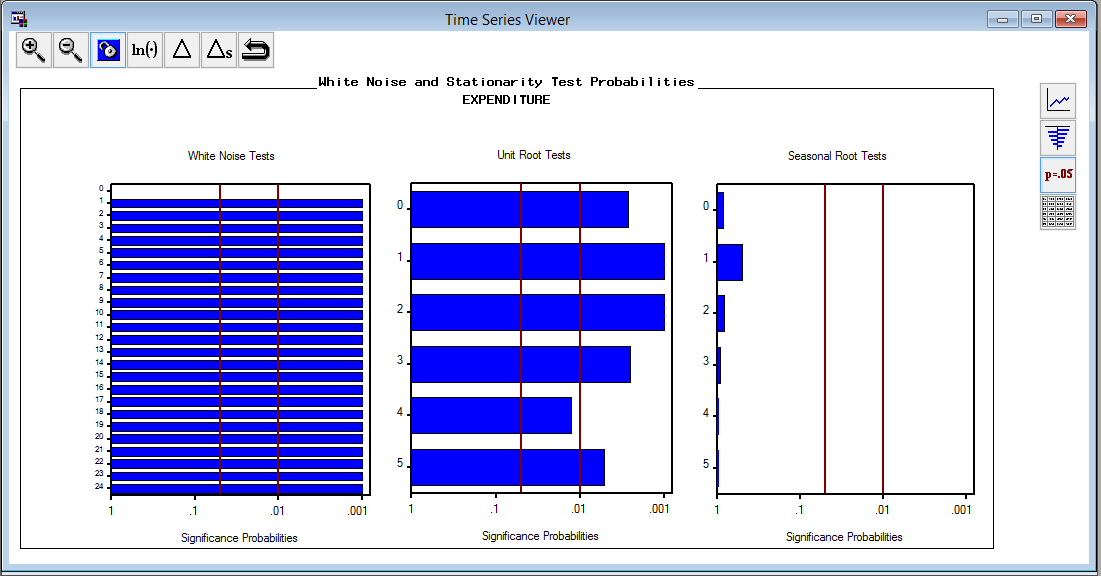




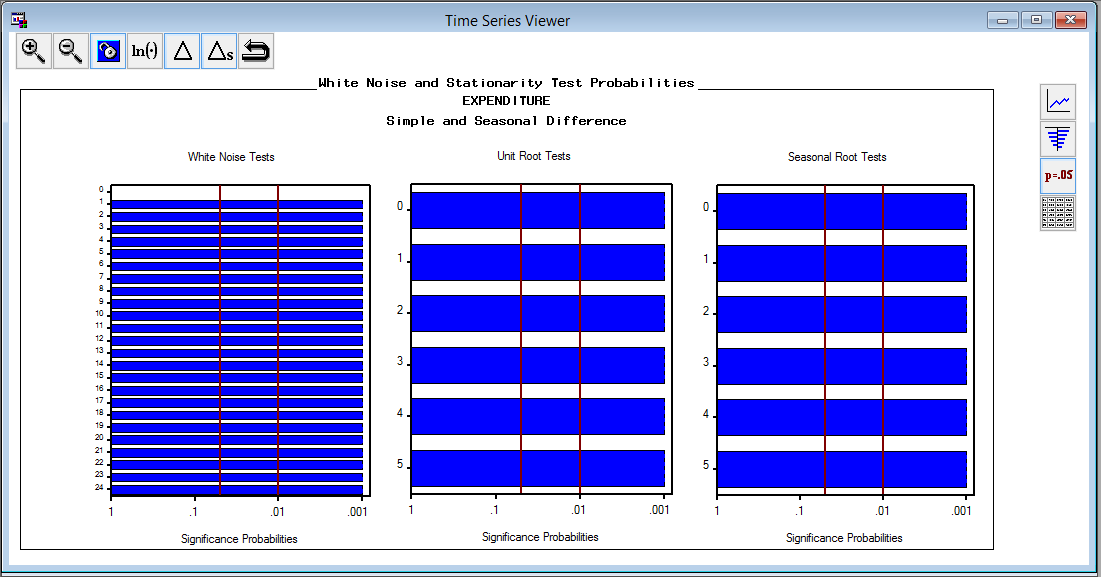
**Monthly Outbound expenditure from UK:**







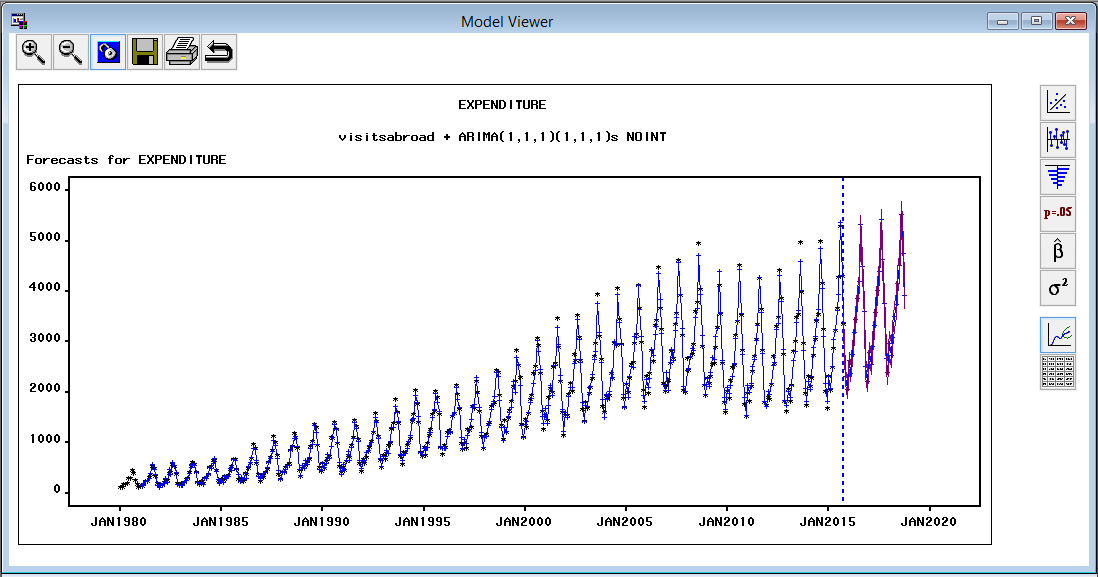


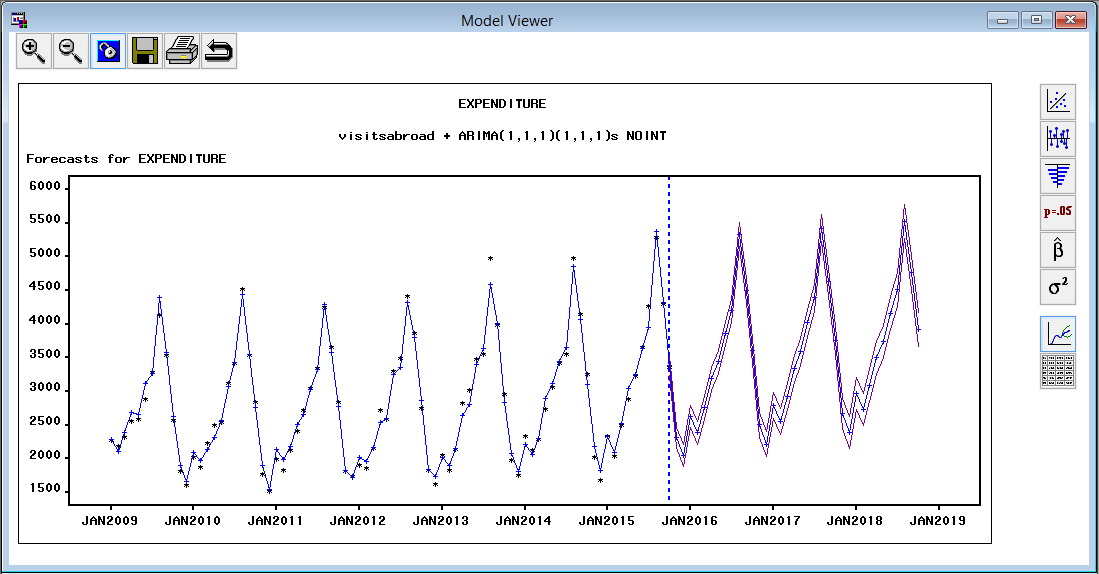


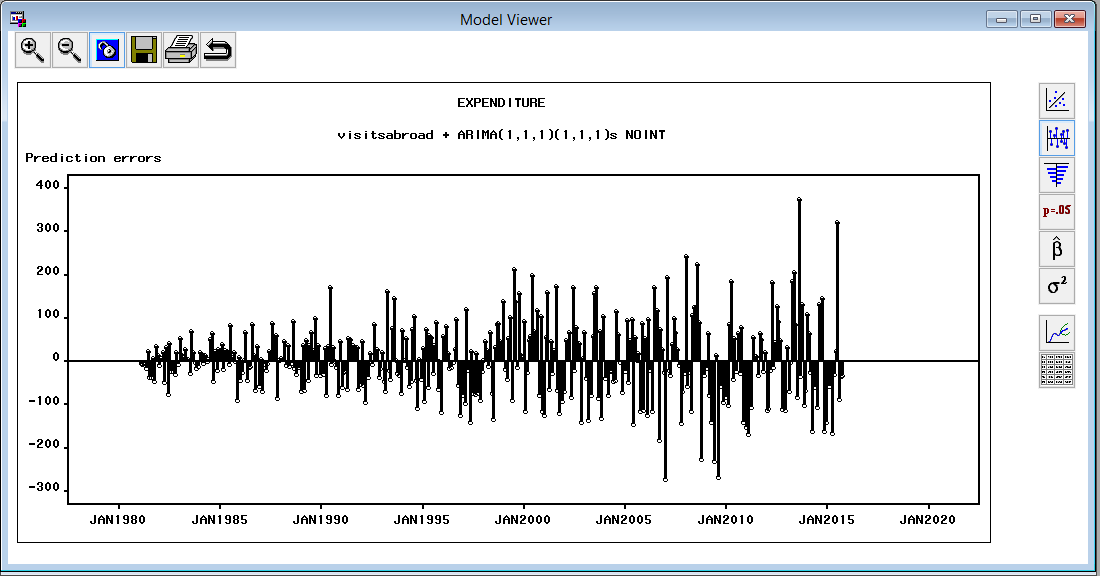
**Candidate Models for Monthly Outbound Expenditure from UK:**

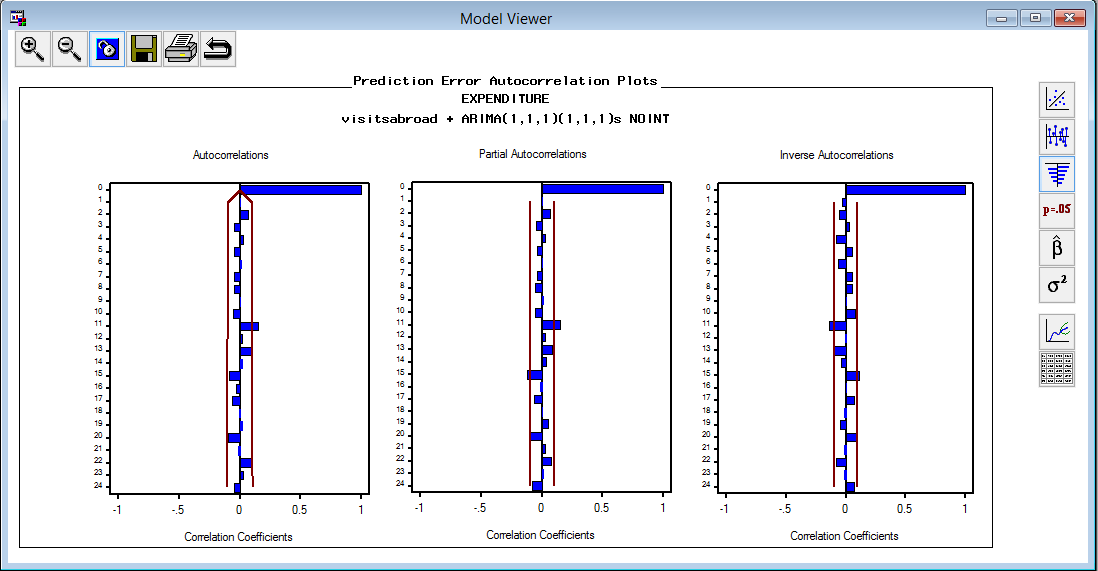


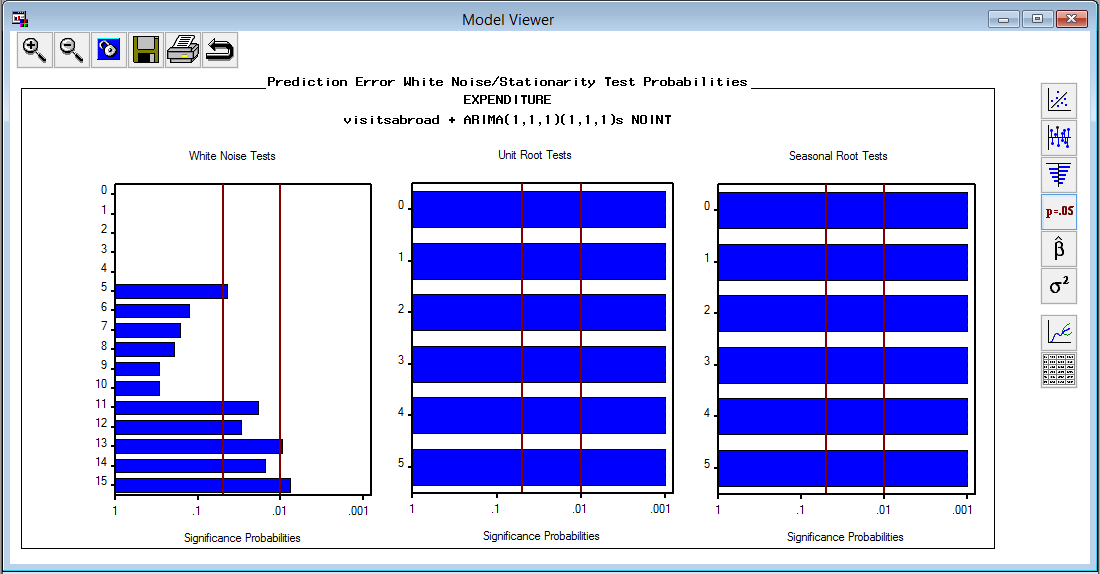
**Forecasted Monthly Outbound expenditure from UK:**

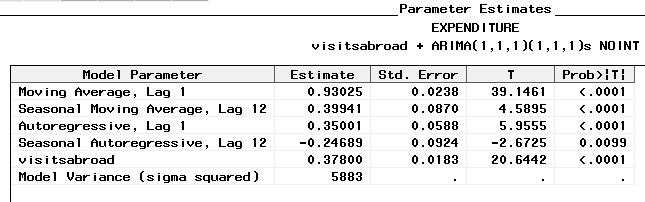


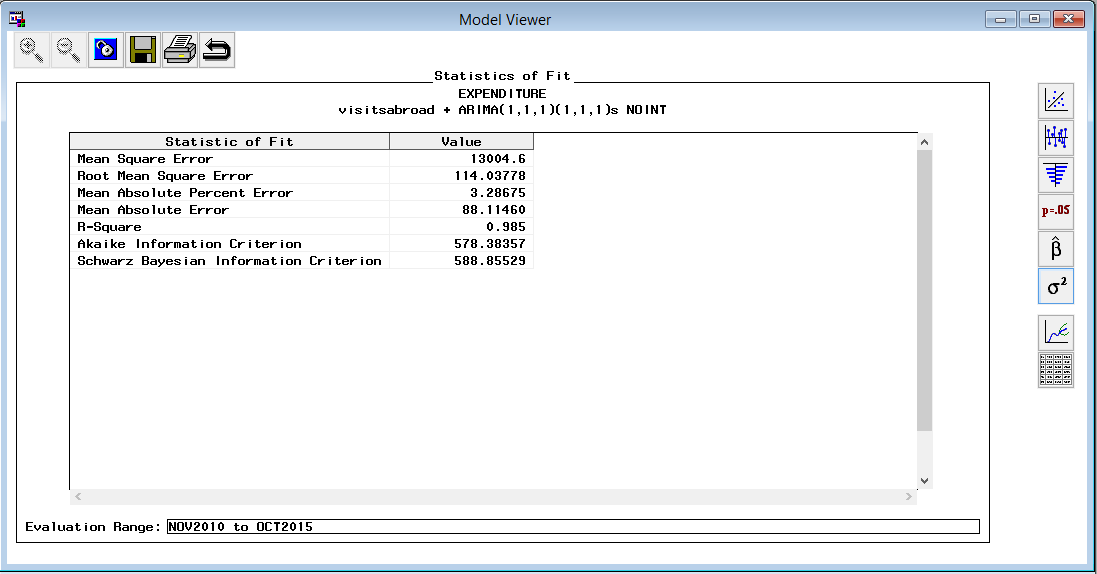






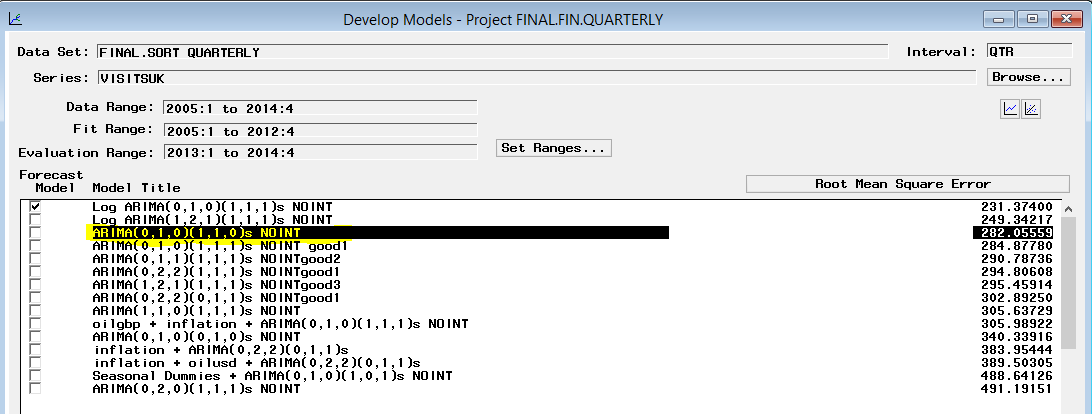




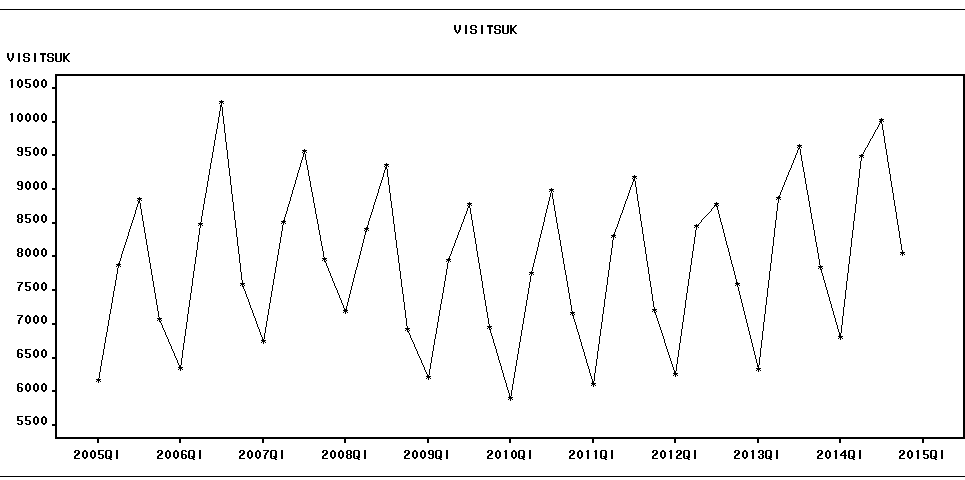


**Quarterly Inbound Visits to UK Forecasts:**

For Quarterly data from 2005 Q1 till 2014 Q4, a holdout sample for 2 years is considered. Having the target variable as ***VisitsUK***, the following candidate models were fit to the series. Out of these models ARIMA(0,1,0)(1,1,0) is selected as the best fitting model after examining ACF,PACF and IACF plots etc.



The following graph indicates the presence of trend and seasonality in the series of data.

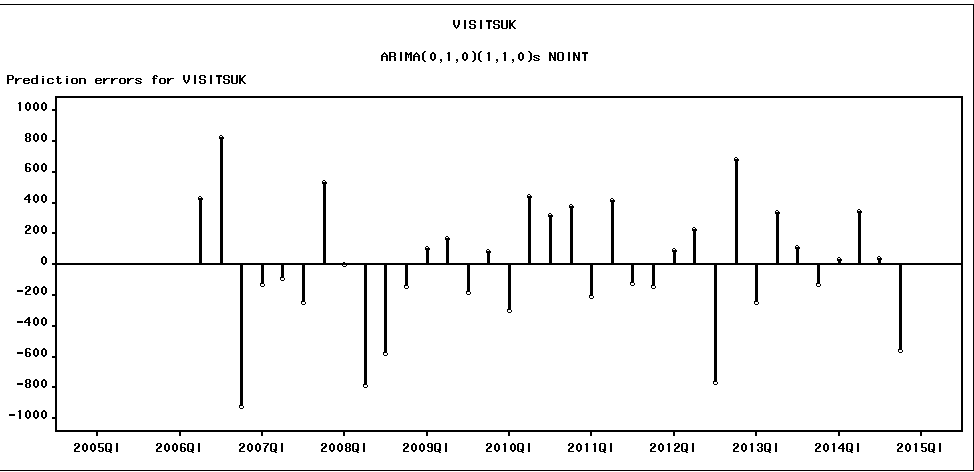


The models were fit to treat the trend and seasonality.

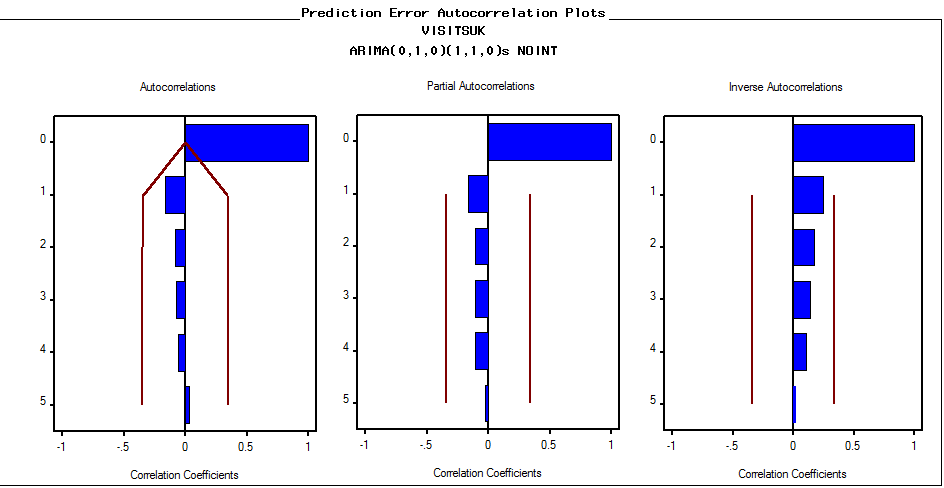
Though the following models ARIMA(0,1,0)(1,1,1), ARIMA(0,1,1)(1,1,1) and ARIMA(1,2,1)(1,1,1) yielded better RMSE, they were rejected as the parameter estimates were not significantly different from zero and as the series is not stationary.

Analysis of the best model: ARIMA (0,1,0) (1,1,0):

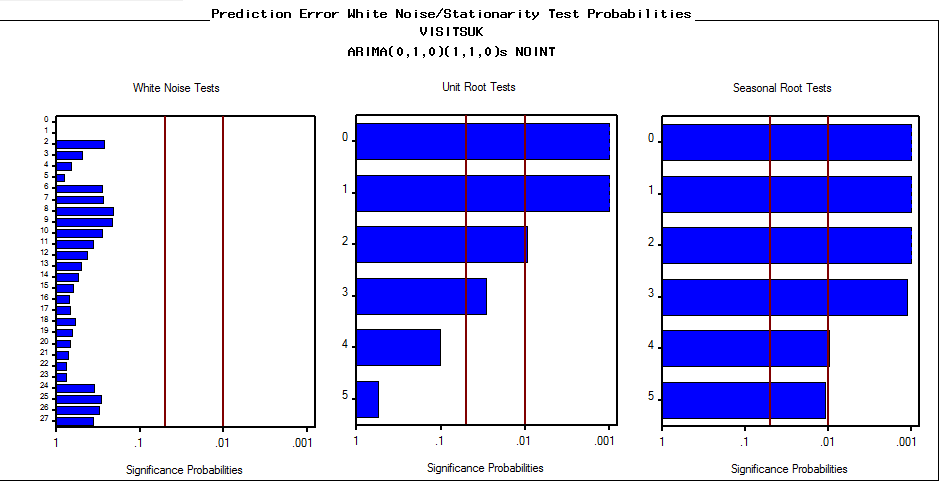
The residuals do not have long run of negative residuals and they seem to be uncorrelated.



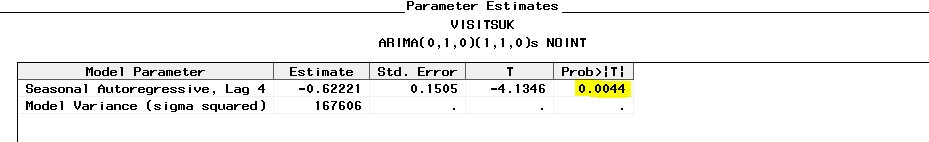
The auto correlations plot indicates that the residuals are uncorrelated



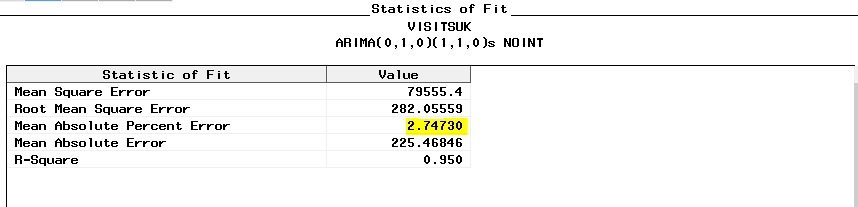
The residuals appear to be white noise and even the series is stationary



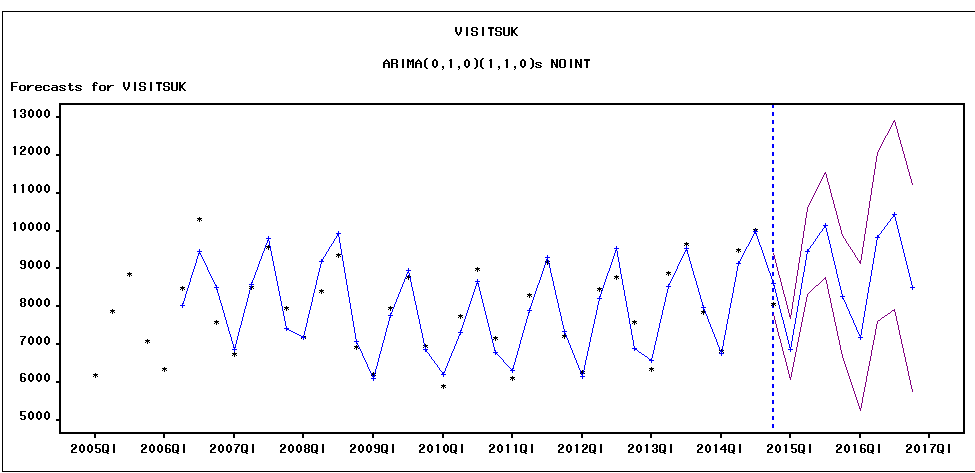
The parameter estimates are significantly different from zero, i.e. significant at 1% level



Accuracy value is about 100-MAPE which seems satisfactory.



The forecasts looks reasonable and the forecasts are not generated with high confidence intervals



**Check for Cross Correlations:**

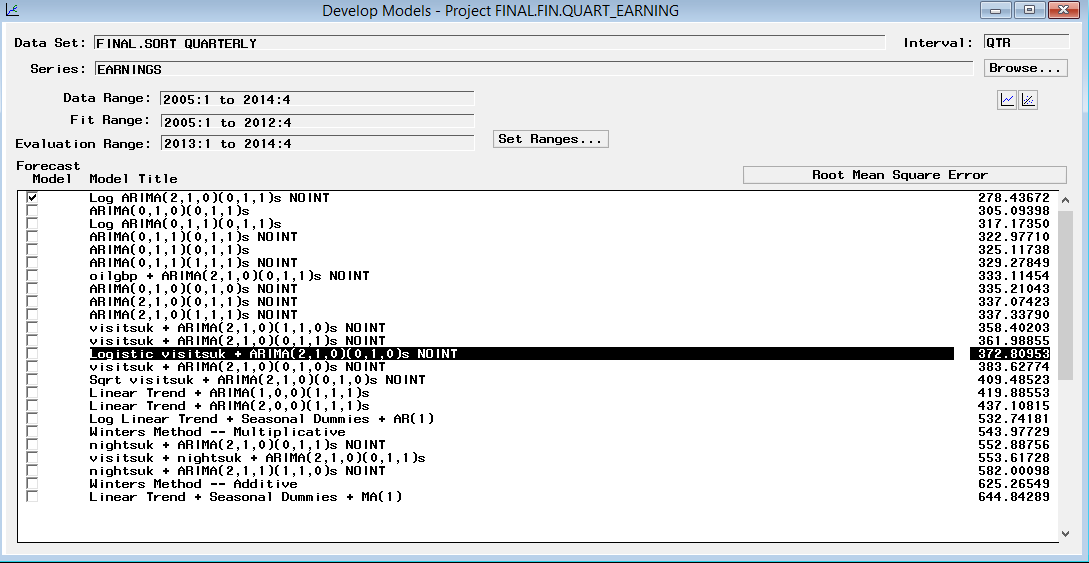
The cross-correlation function is computed after any specified differencing has been done. If differencing is specified for the VAR= variable or for a variable in the CROSSCORR= list, it is the differenced series that is cross-correlated. Compute the cross-correlations of the changes in Y with the changes in X. The cross correlated variables will be used as dynamic repressors.

The code to check for cross correlations for the above model is as follows:

****

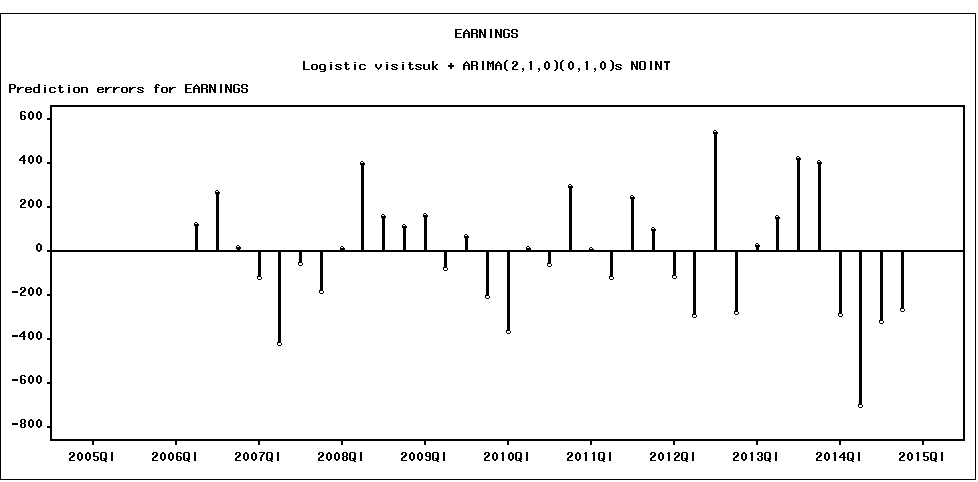
**Quarterly inbound Earnings to UK Forecasts:**

Considering the target variable as ***Earnings***, the following candidate models were fit to the series.



Analysis of the best model based on RMSE criteria: logistic visitsuk +ARIMA (2,1,0)(0,1,0)

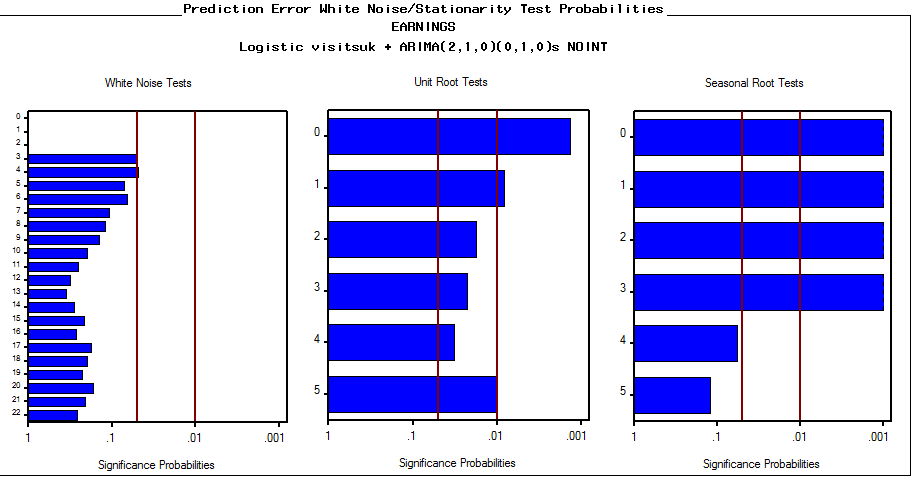
There are no long runs in the residuals. It seems proper



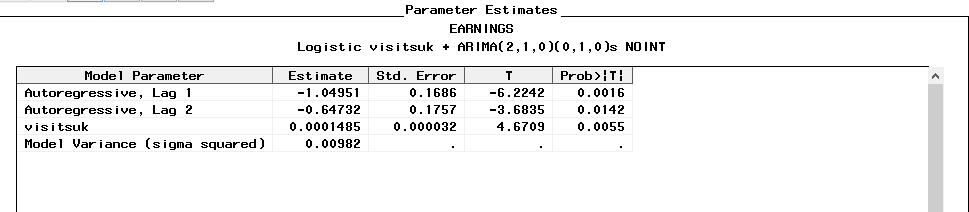
The auto correlation plot suggests that there are no significant auto correlations in the plot



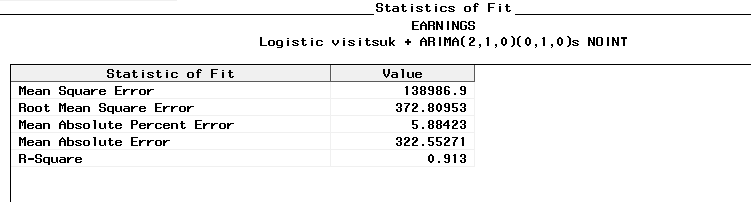
The data passes both the white noise and unit root tests indicating that the errors are randomly distributed



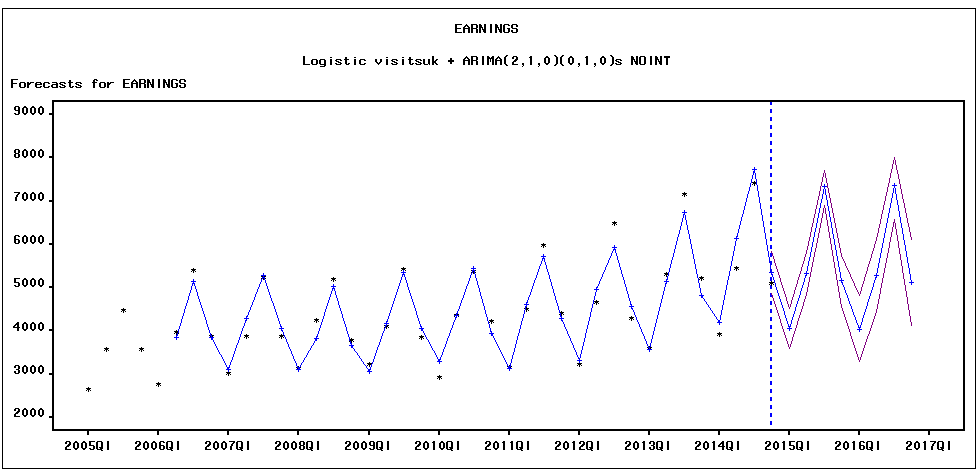
The parameter estimates are significantly different from zero



MAPE value is about \*% accuracy which seems satisfactory.



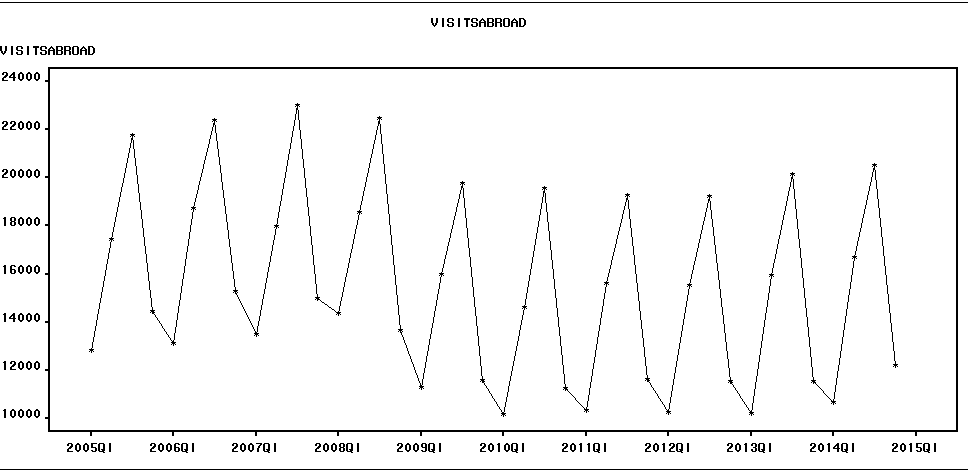
The forecasts looks reasonable and the forecasts are not generated with high confidence intervals

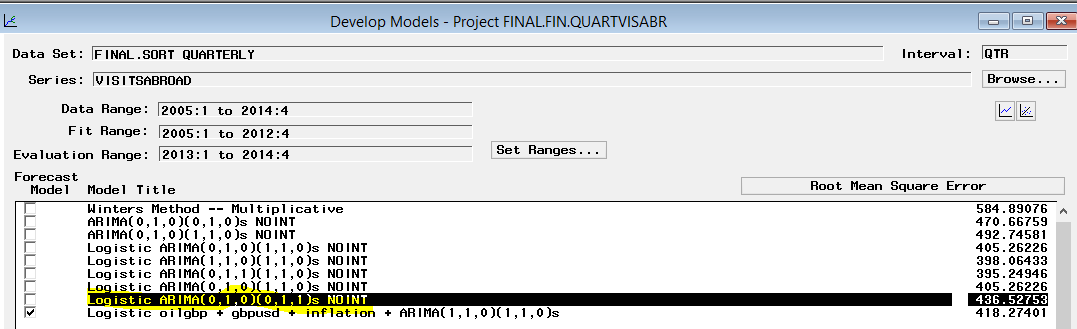


**Visits Abroad Forecasts**

Considering the target variable as ***Visitsabroad***, the following candidate models were fit to the series.

From the graph series it is observed that there is trend and seasonal components and also there is no unusual trend at the end of series indicating that the hold out sample can be taken for this data

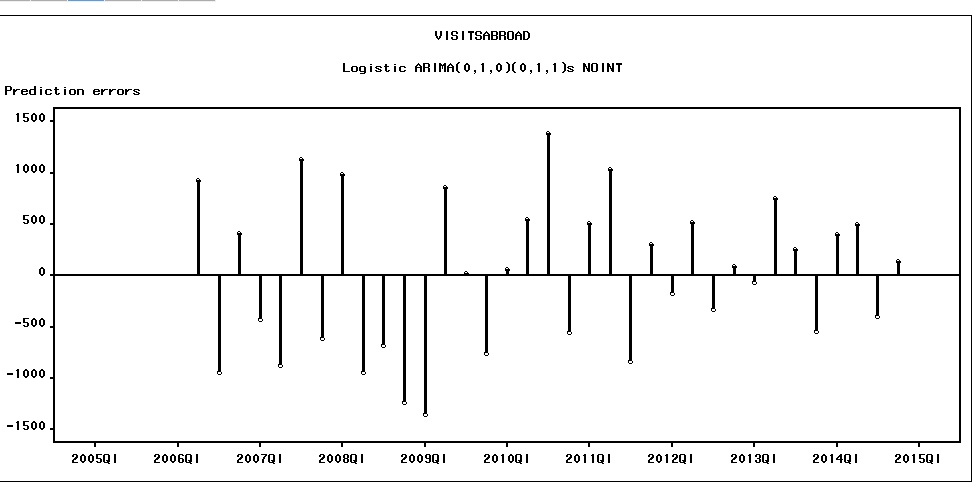




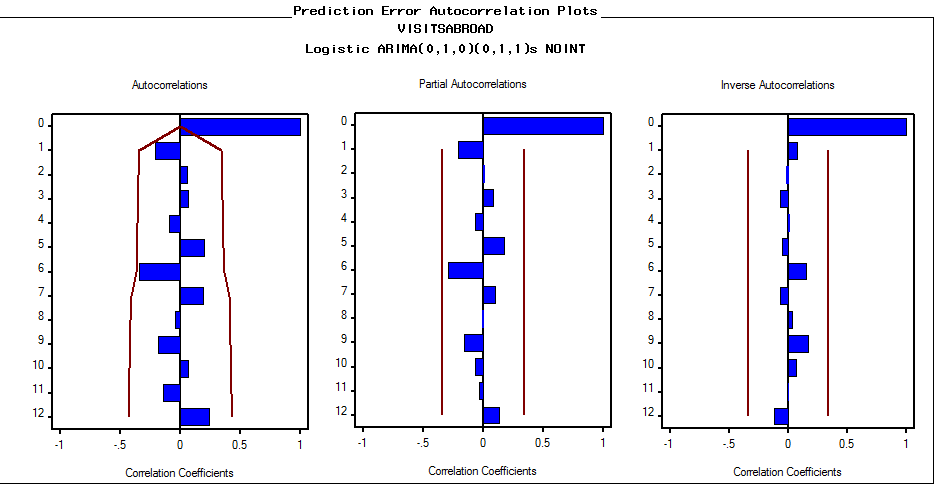
After applying first and seasonal difference the error component is stationary but the white noise clearly indicate the presence of autocorrelation. Therefore need to build the ARIMA models to investigate the error component

Analysis of the best model based on RMSE criteria: logistic ARIMA(0,1,0)(0,1,1)

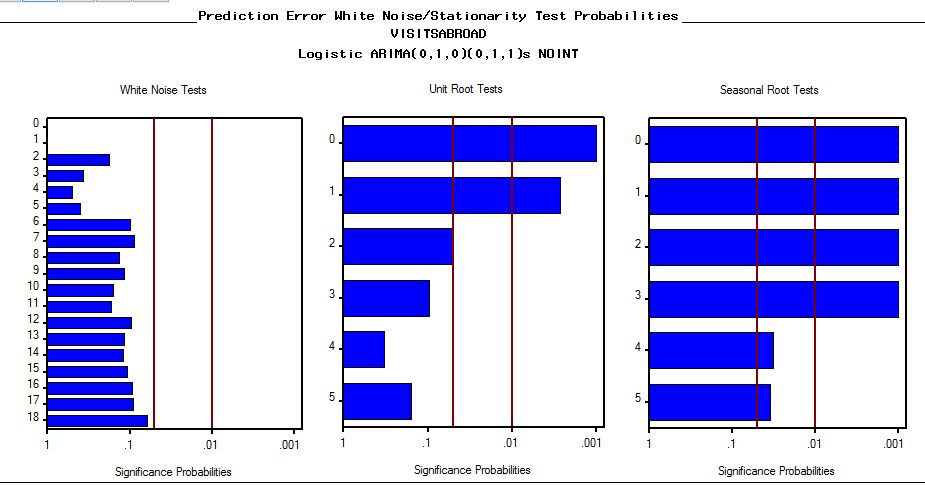
1. Residual Analysis: There are no long runs in the residuals. It seems proper



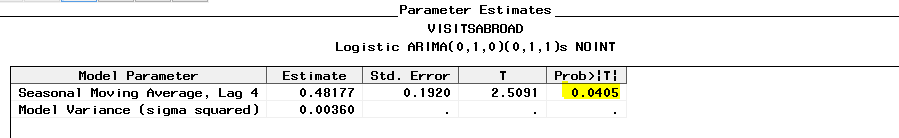
The residuals appear to be random. There are no significant auto correlations in the ACF Plots



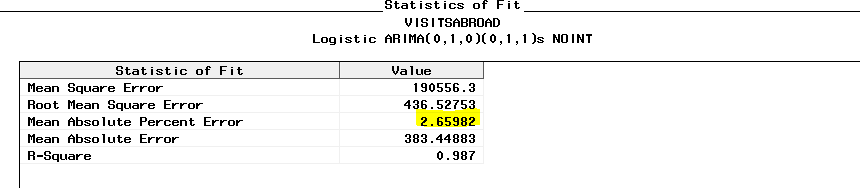
The white noise tests imply uncorrelated residuals. Unit root tests indicate stationarity, at least at the 5% level of significance.



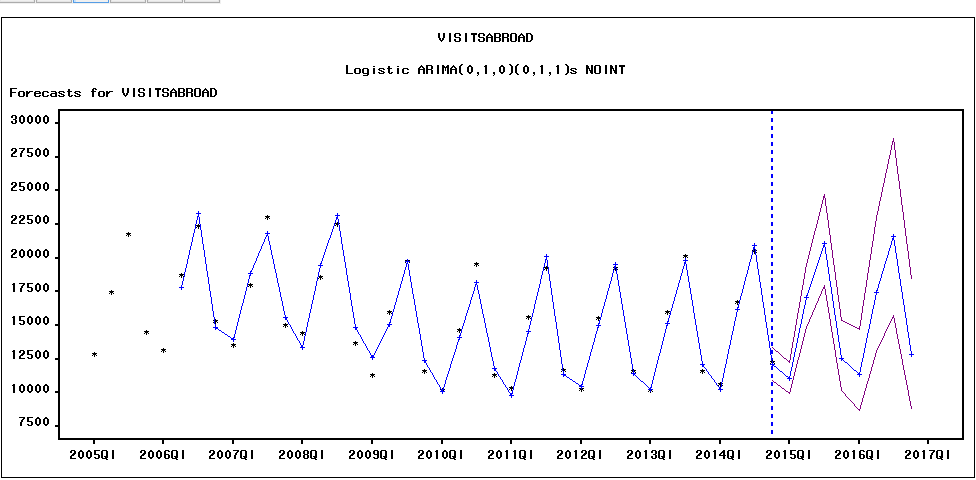
The parameter estimates are significantly different from zero



MAPE indicates an average error of about 2.65%. The evaluation range reminds you that the coefficient was estimated using the entire time series, not just the first 8 periods.



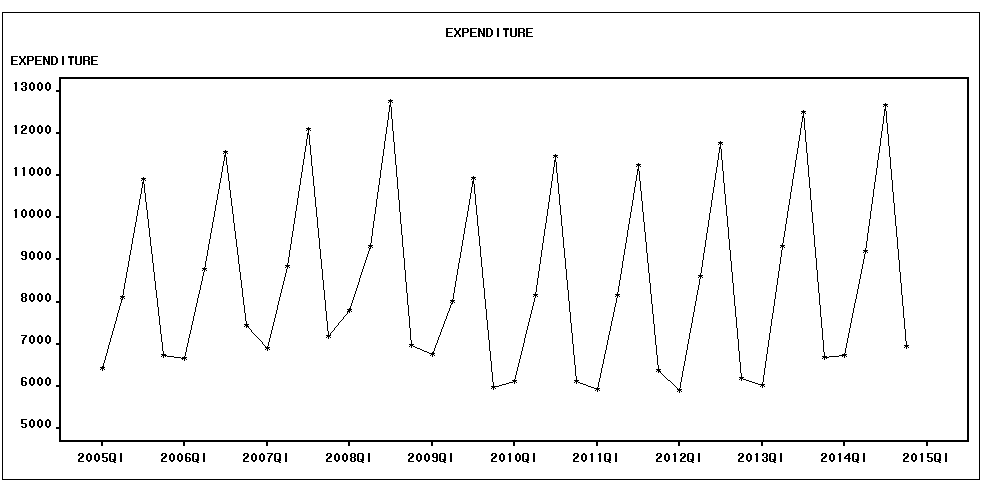
The forecast plot looks fine and the presence of trend and seasonality produces tight 95% prediction limits

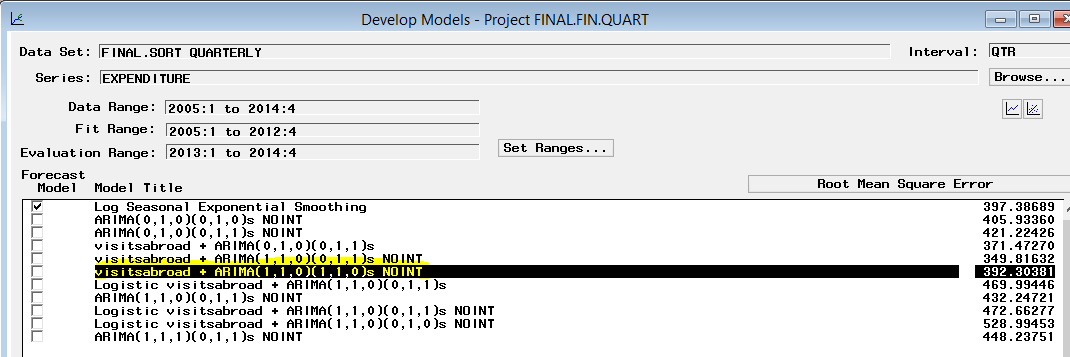


**Quarterly Outbound Expenditure Forecasts:**

Considering the target variable as ***Expenditure***, the following candidate models were fit to the series.

From the graph series it is observed that there is trend and seasonal components and also there is no unusual trend at the end of series indicating that the hold out sample can be taken for this data

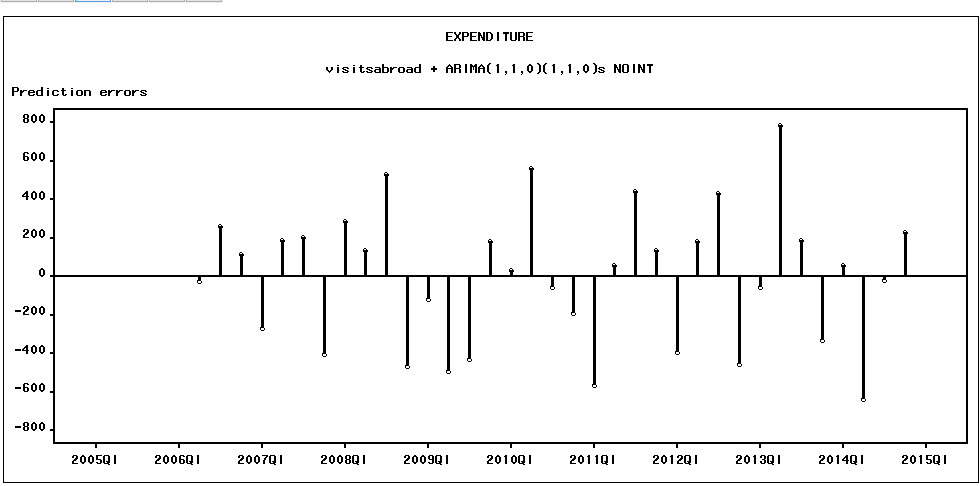




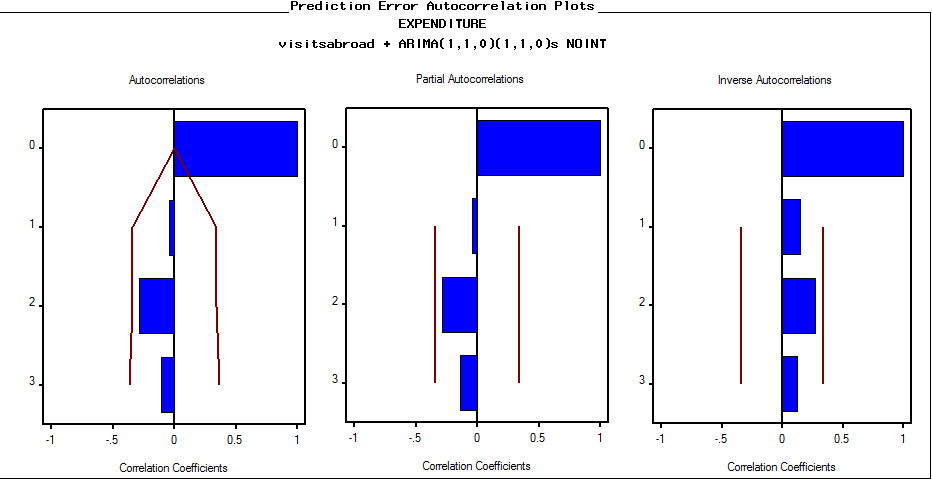
After applying first and seasonal difference the error component is stationary but the white noise clearly indicates the presence of autocorrelation. Therefore, need to build the ARIMA models to investigate the error component

Analysis of the best model based on RMSE criteria: visitsabroad +ARIMA(1,1,0)(1,1,0)

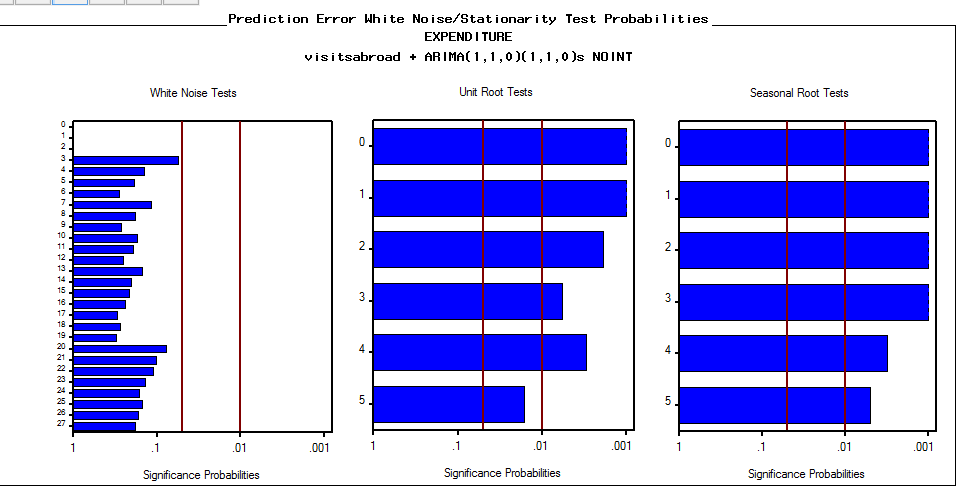
The residual analysis fails to disqualify this model as there no long runs of negative errors and the errors look random



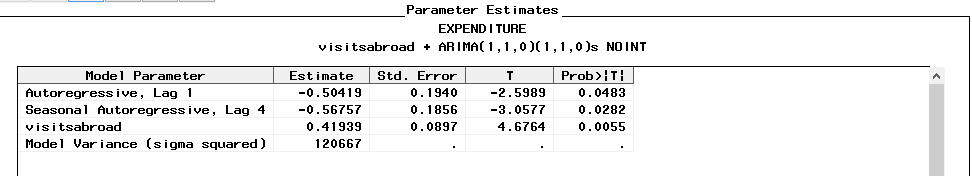
The residuals appear to be random. There are no significant auto correlations in the ACF Plots



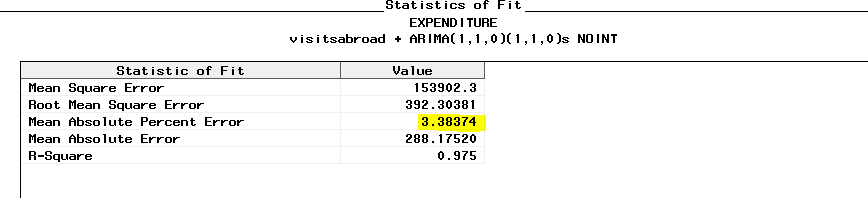
The white noise tests imply uncorrelated residuals. Unit root tests indicate stationarity.



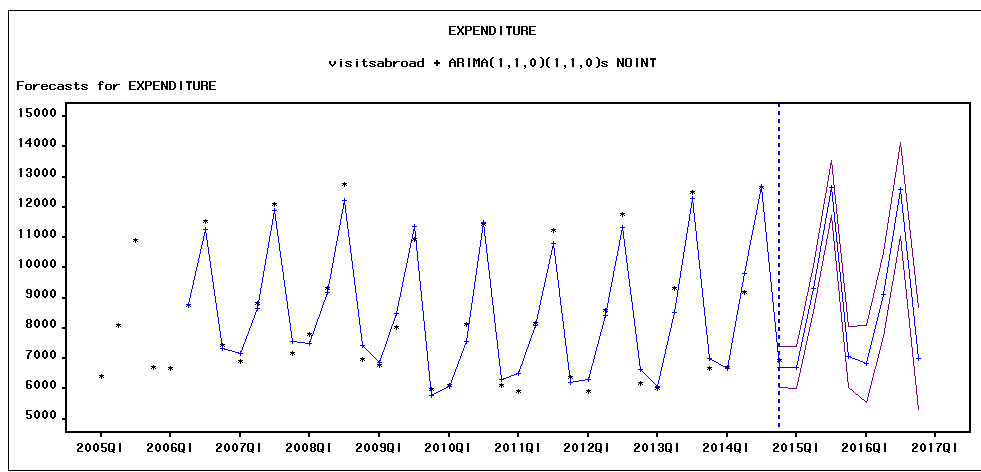
The parameter estimates are significantly different from zero at 5% level. This fact is insufficient to disqualify the model



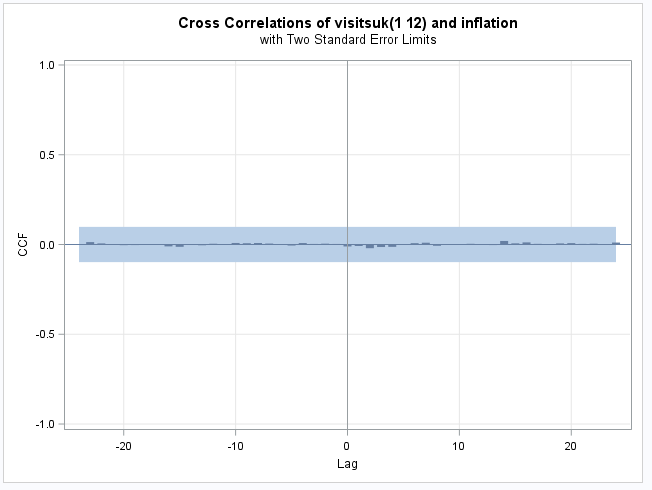
MAPE indicates an average error of about 3.38%. The evaluation range reminds you that the coefficient was estimated using the entire time series, not just the first 8 periods.

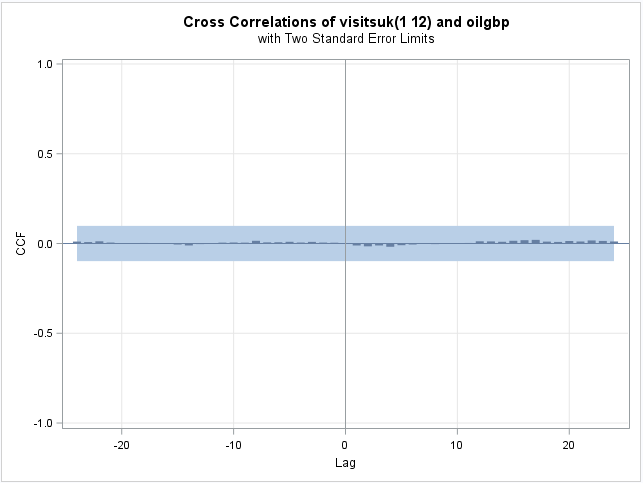


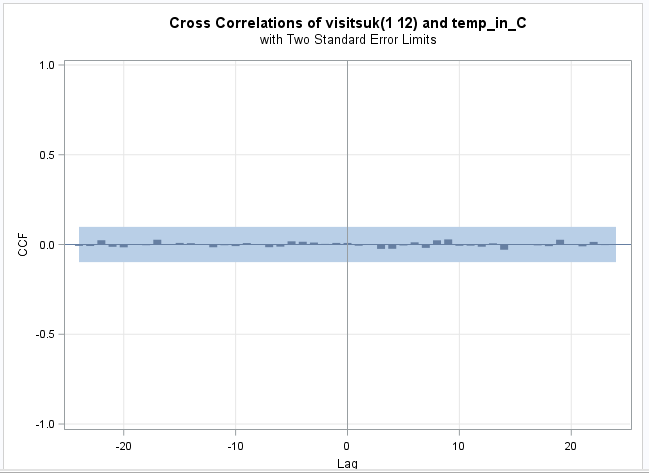
The forecast plot looks fine and the presence of trend and seasonality produces tight 95% prediction limits

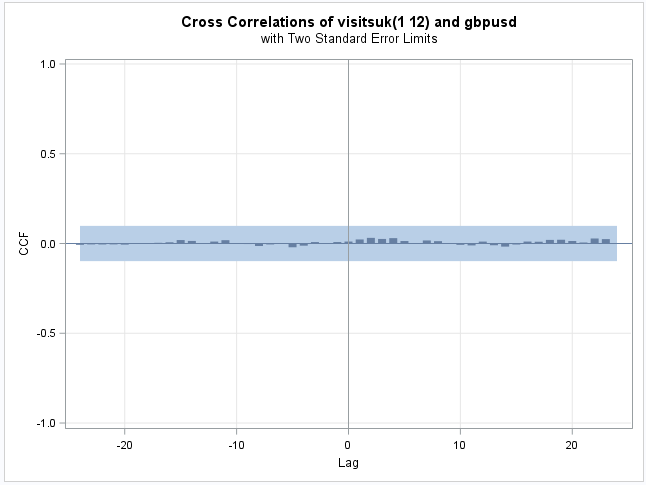


**Cross Correlation graphs:**



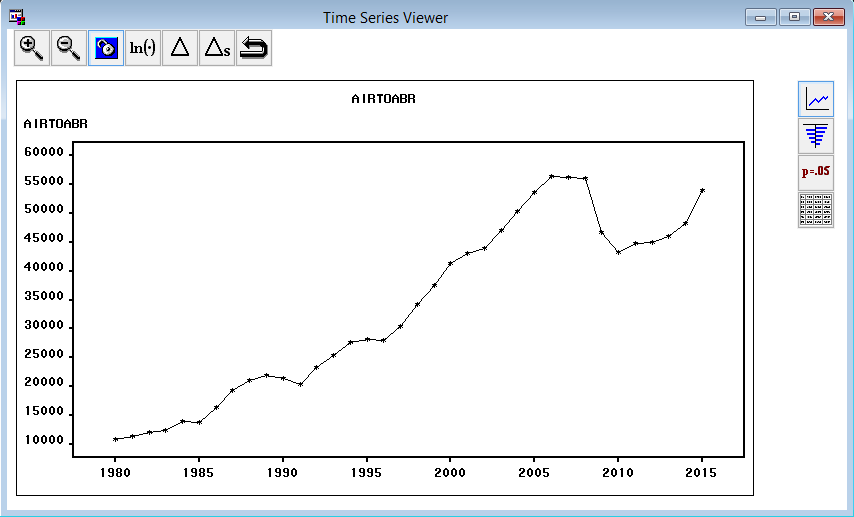




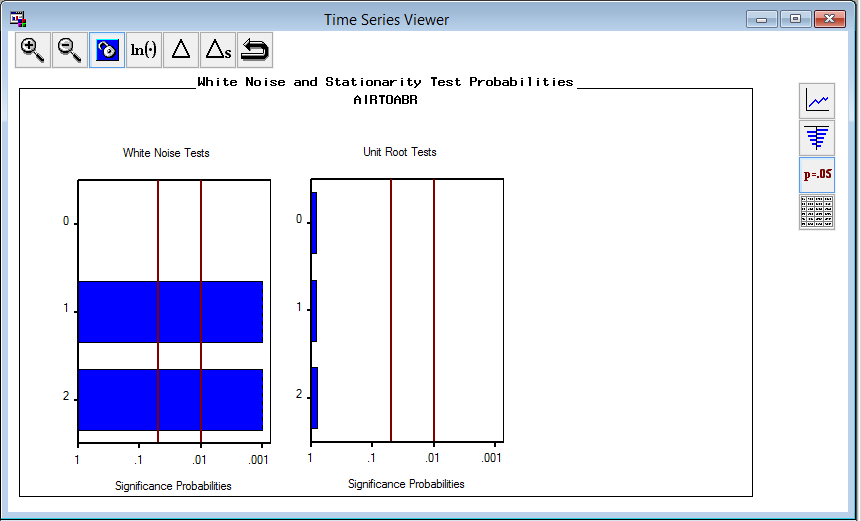


**UK Residents travel abroad by Air**

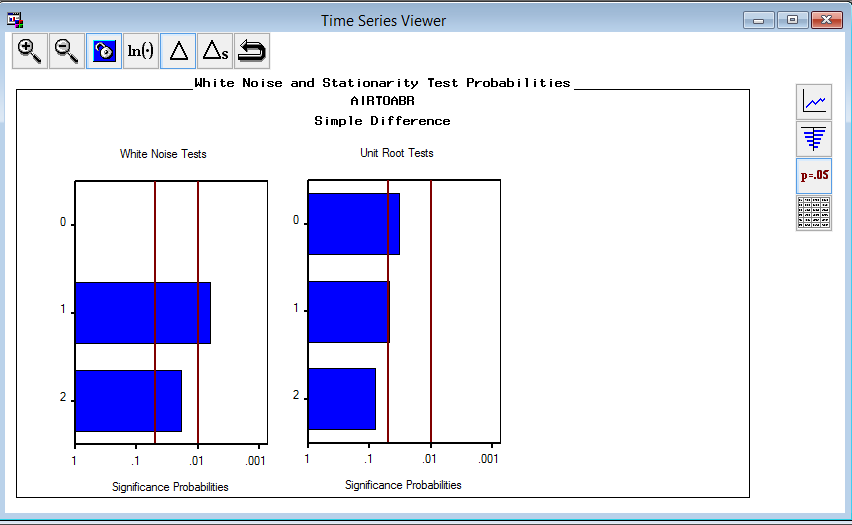
For yearly data from 1980 to 2014;

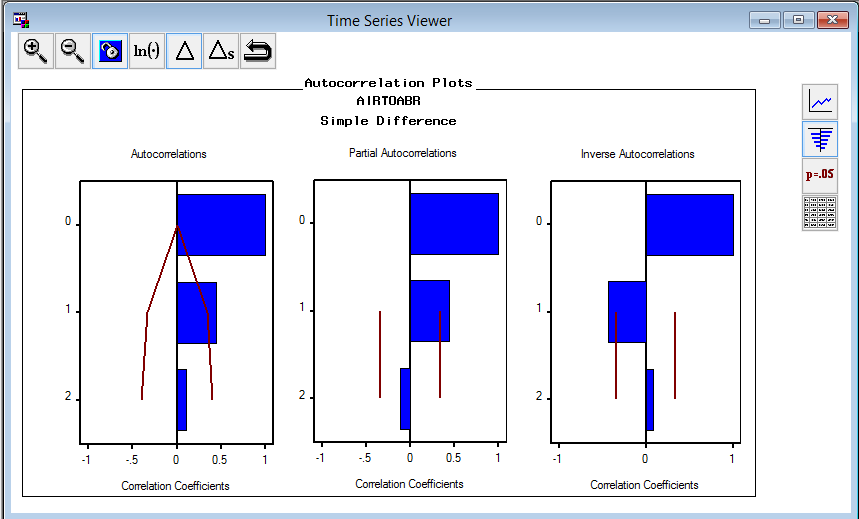






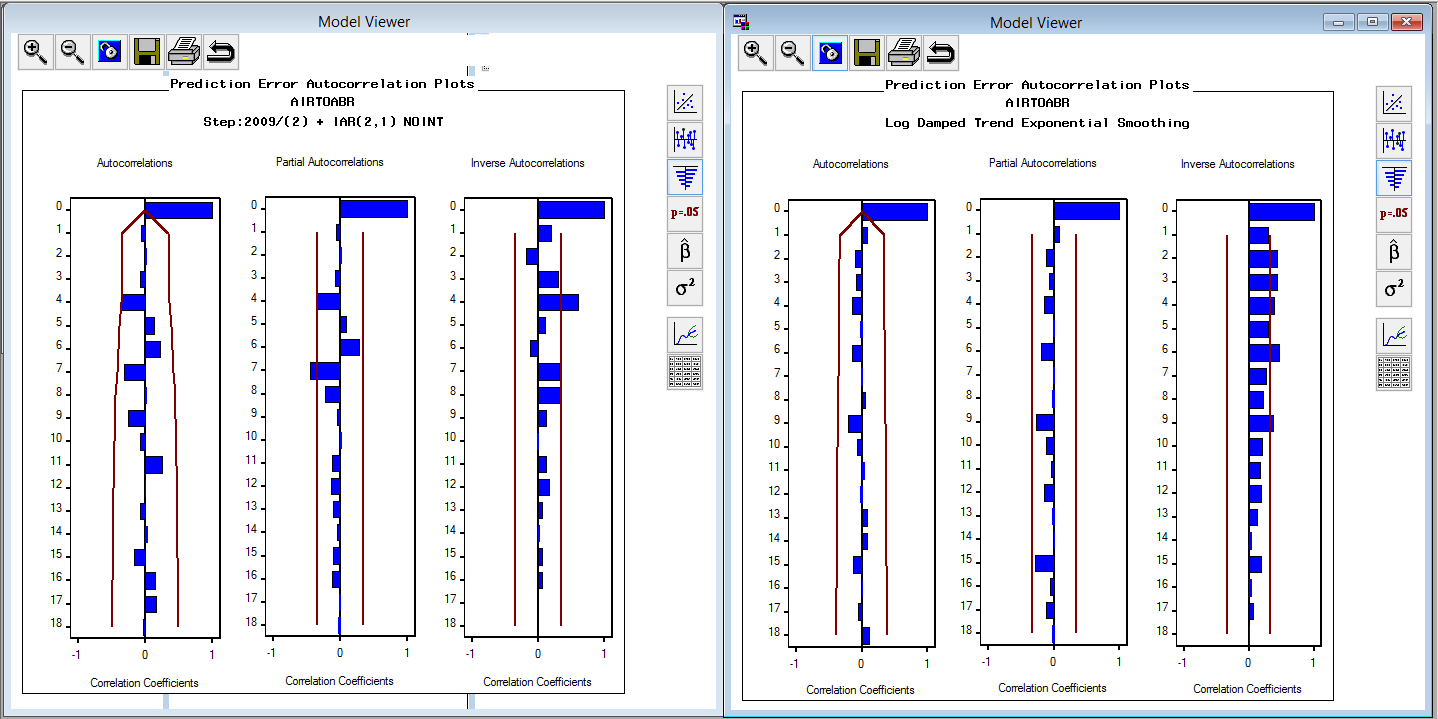
After first differencing, white noise and stationary test improves.

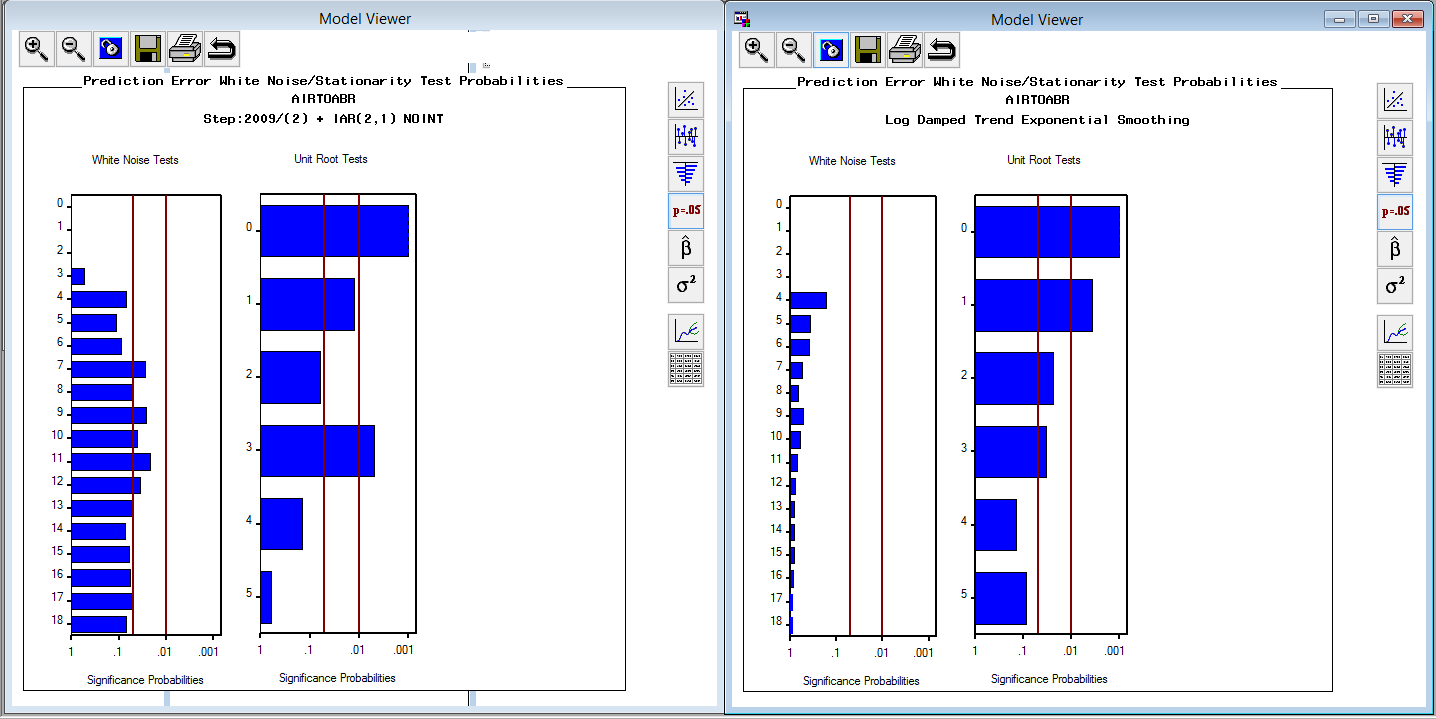


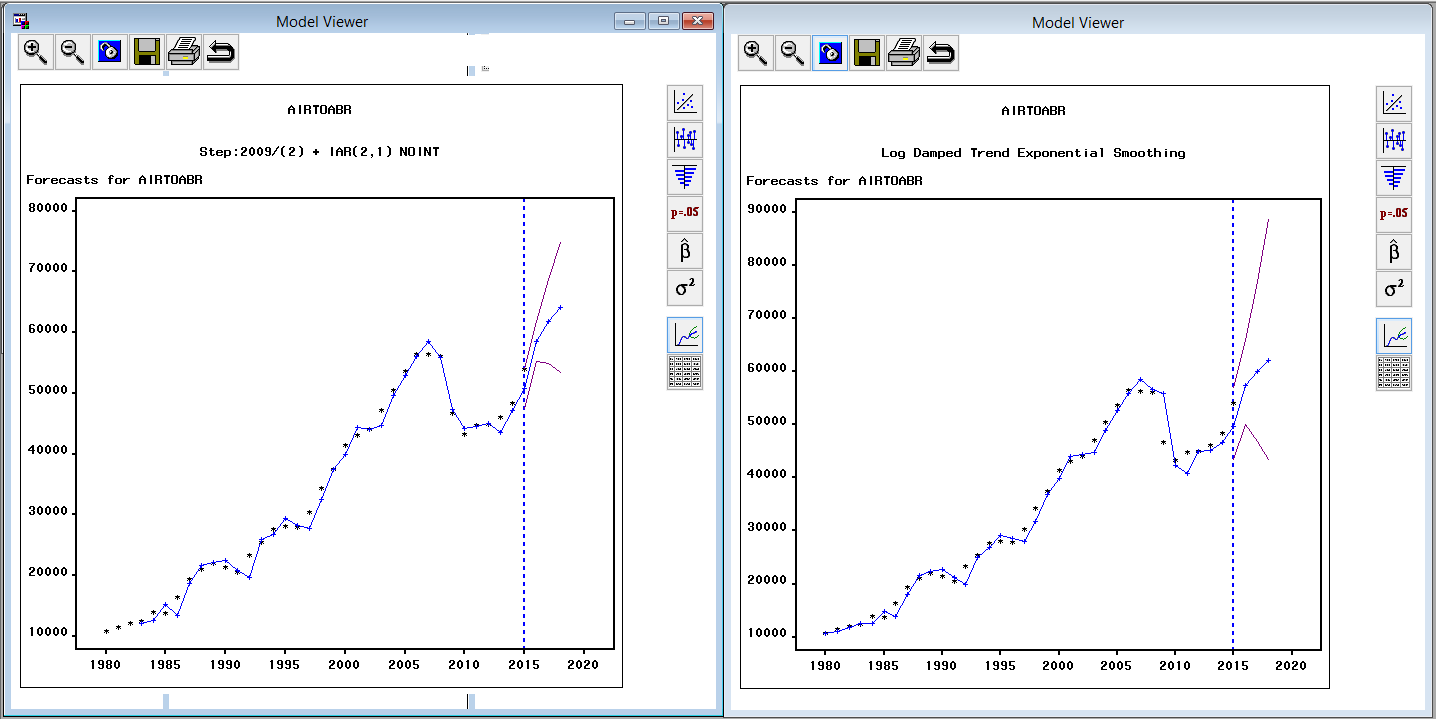


Many model combinations were tried, among which ARIMA(1,1,2) with step intervention (wave) is selected as the best fitting model after examining ACF,PACF and IACF plots etc.



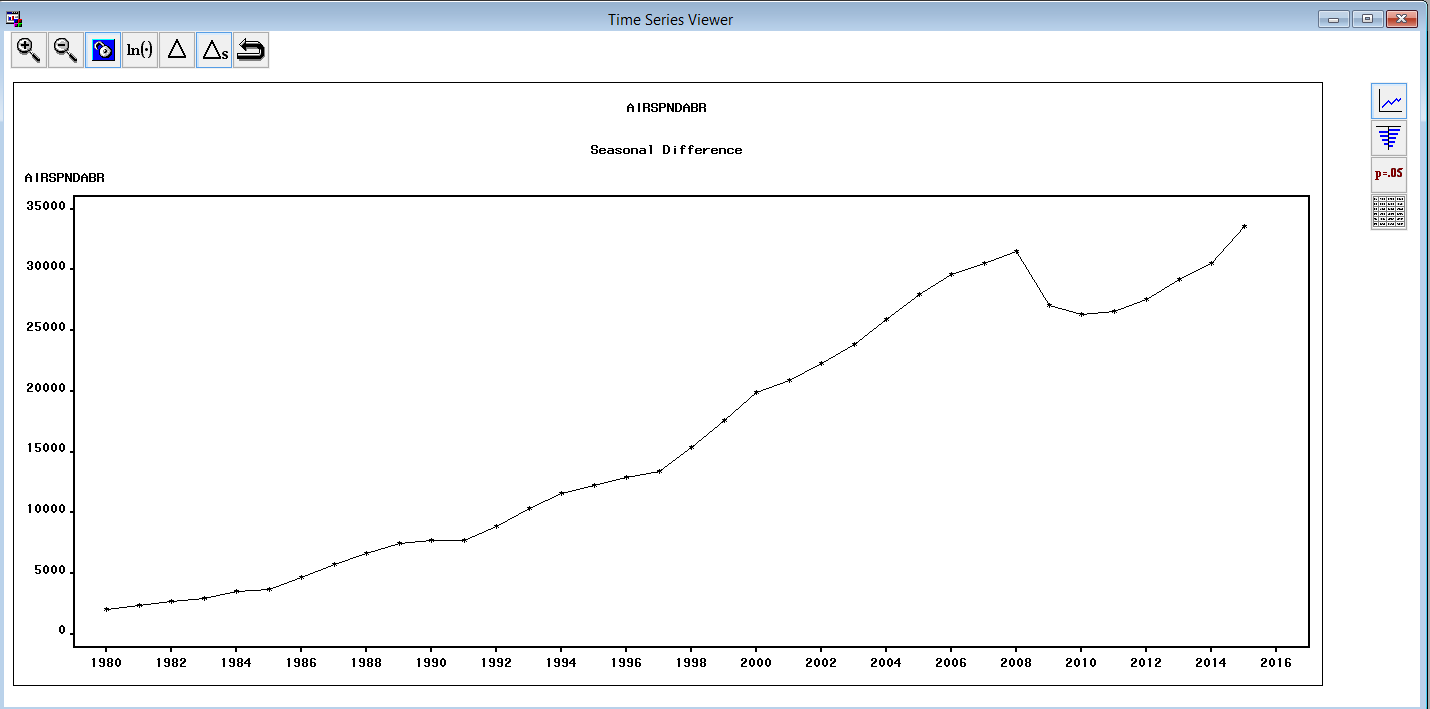


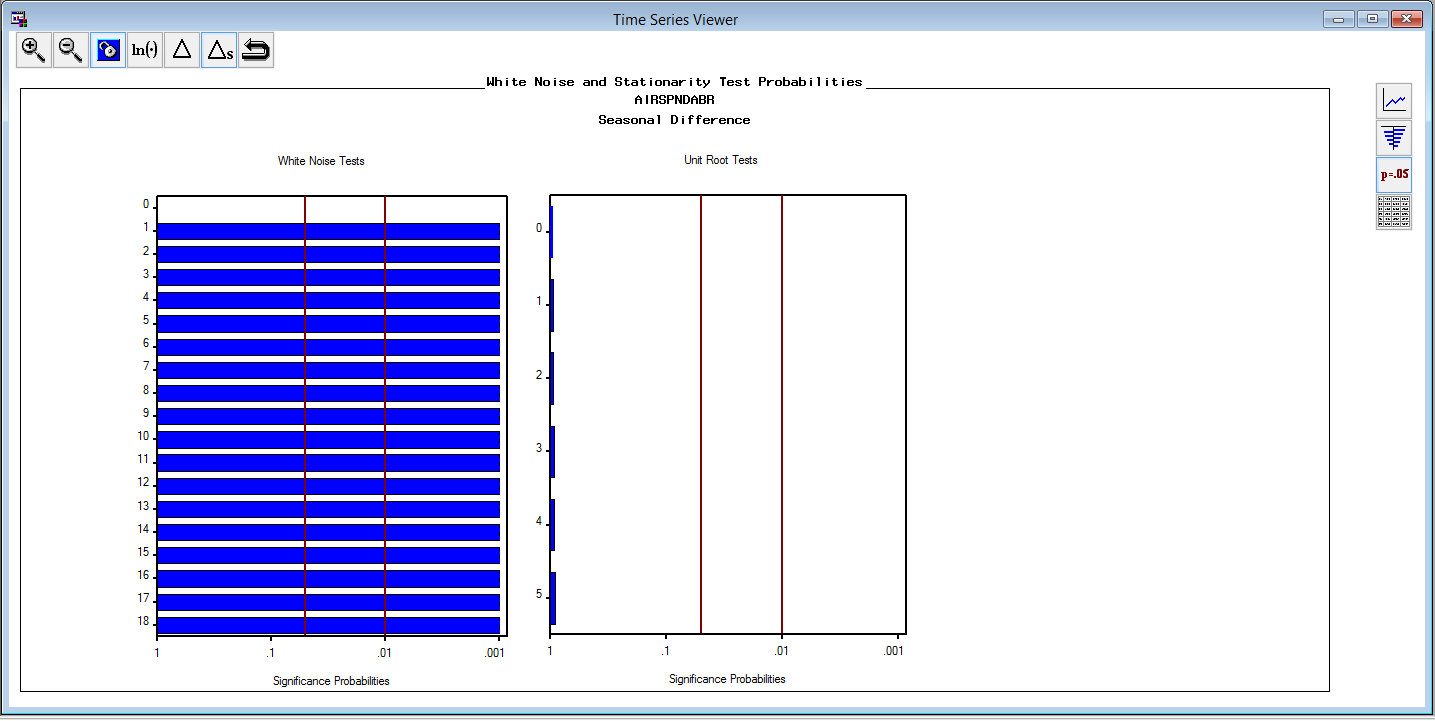


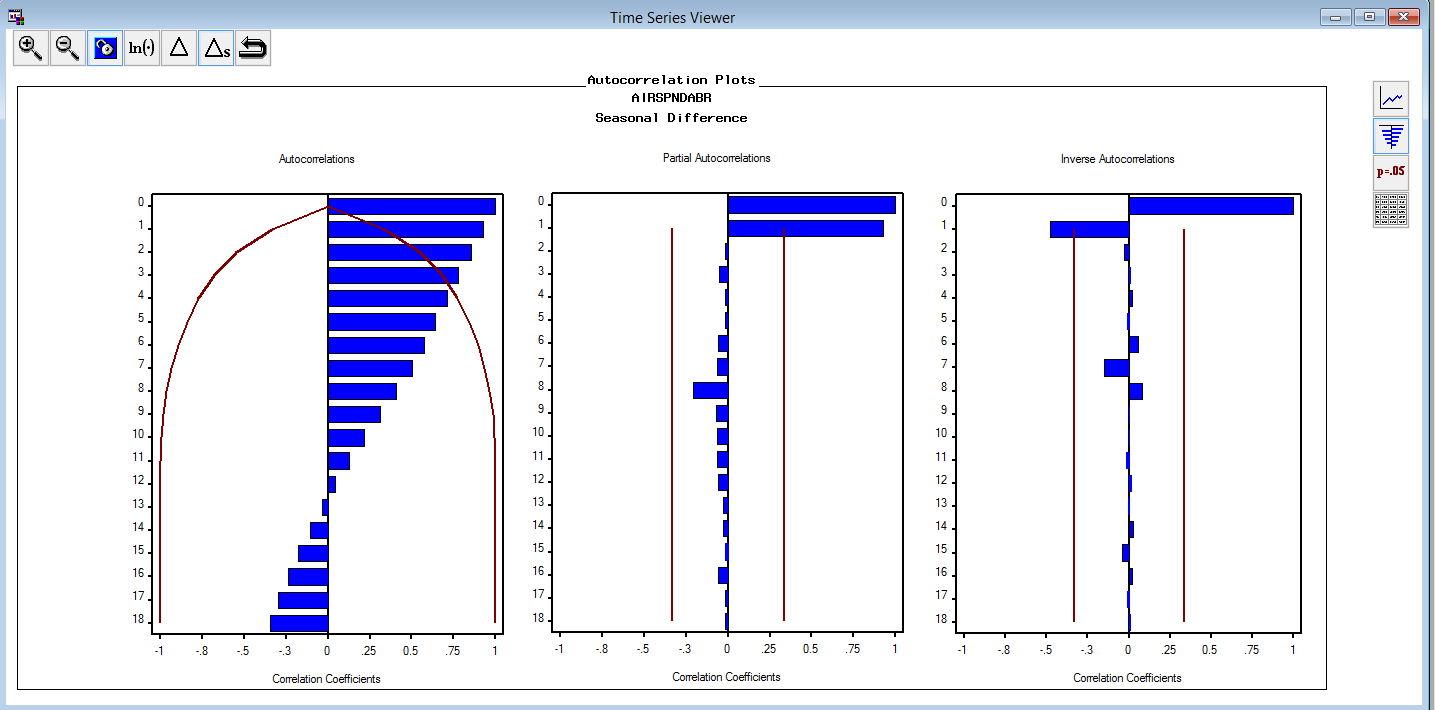


**UK Residents Spending on AIR Abroad**

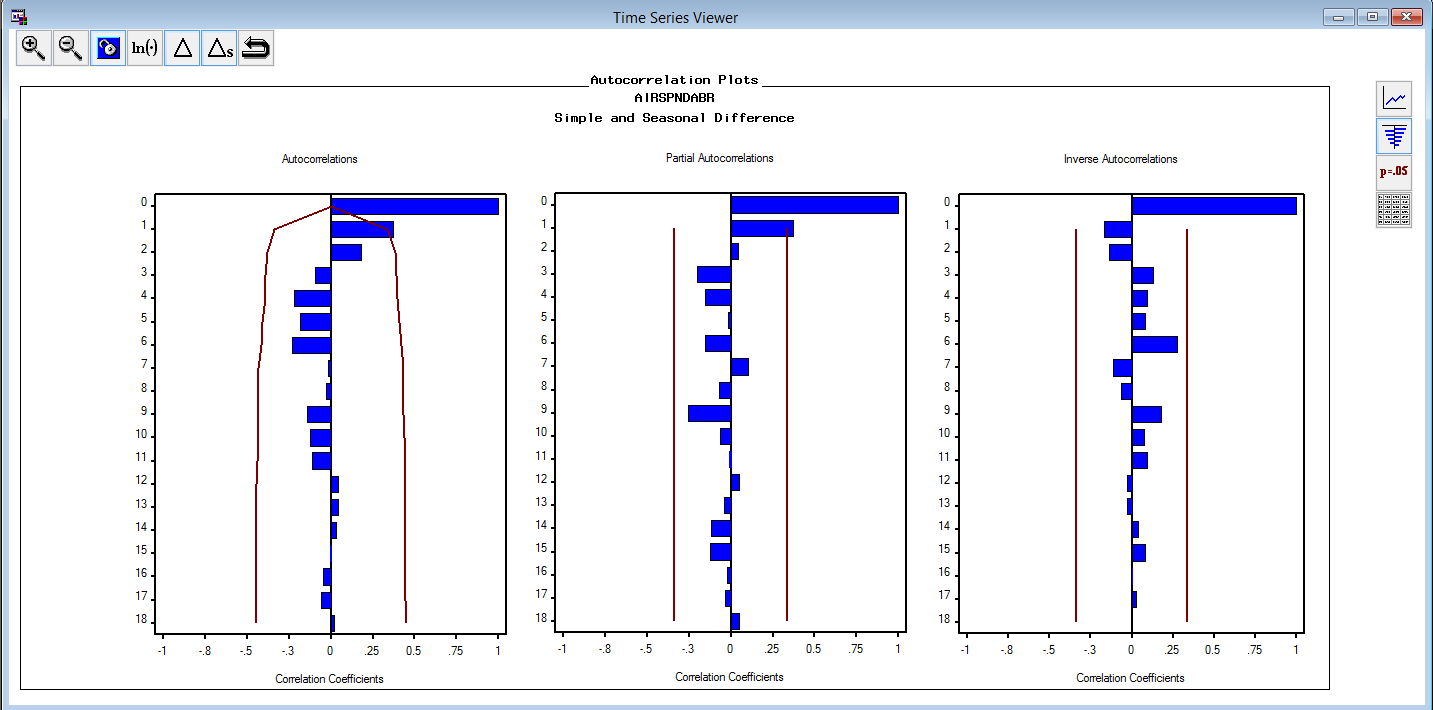
For yearly data from 1980 to 2014;

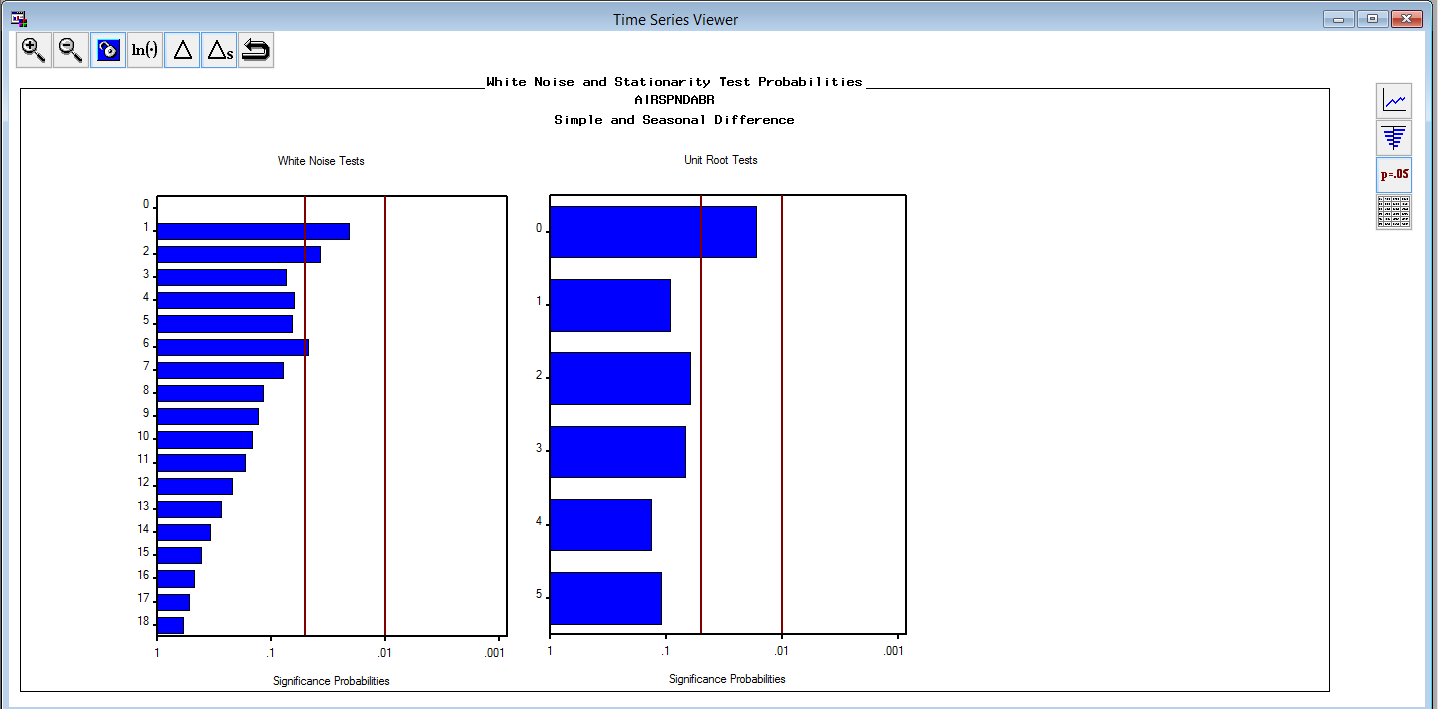




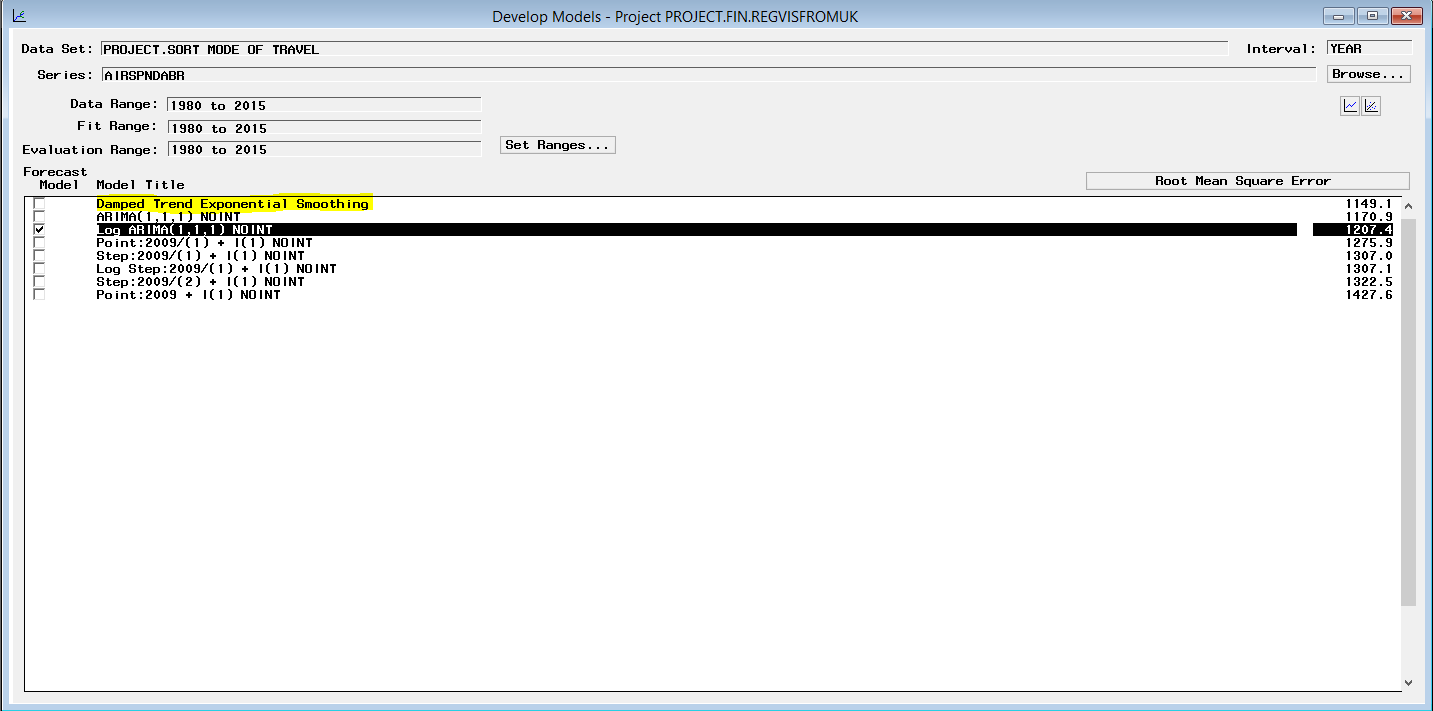


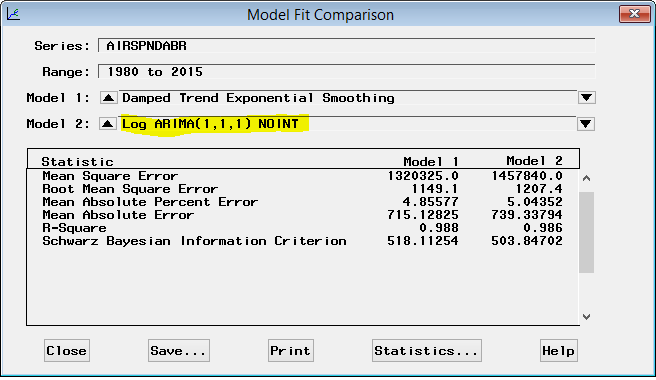
After first differencing,

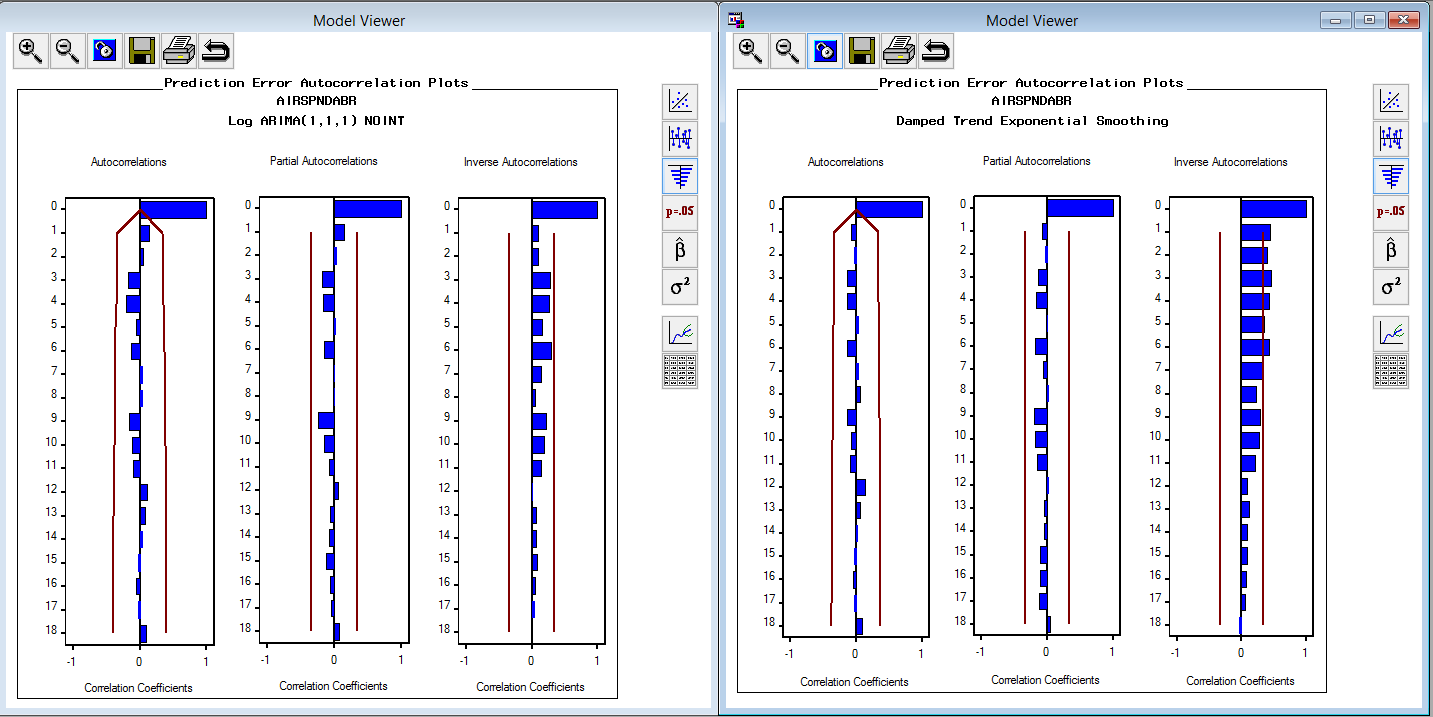


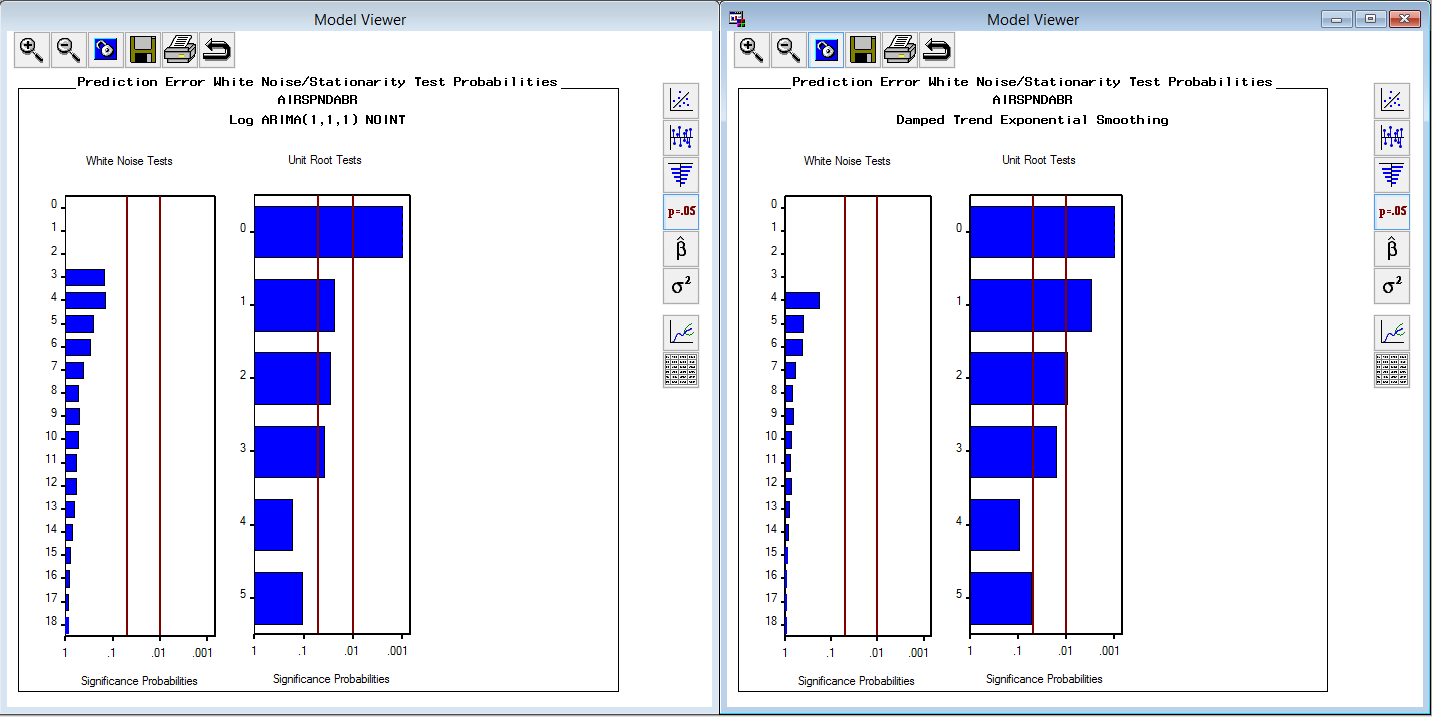


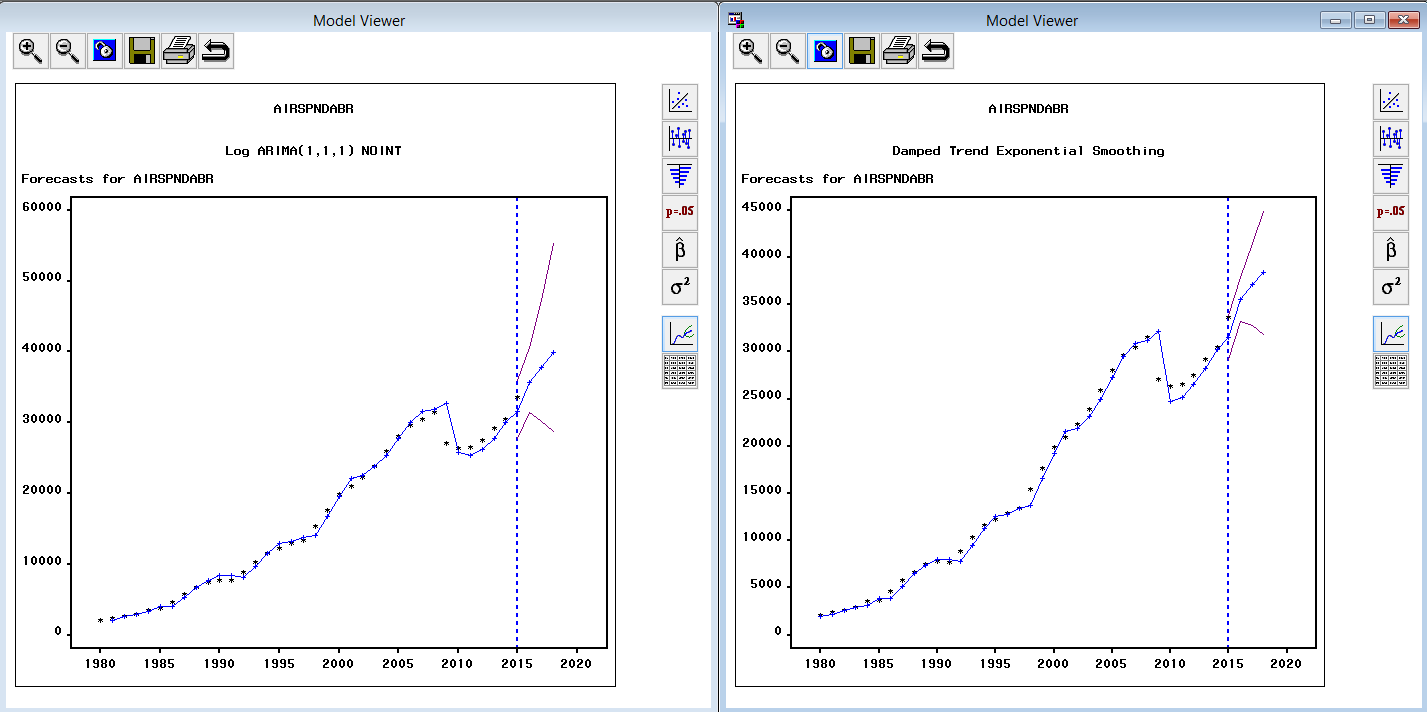
Many model combinations were tried, among which Log ARIMA(1,1,1) is selected as the best fitting model after examining ACF, PACF and IACF plots etc.



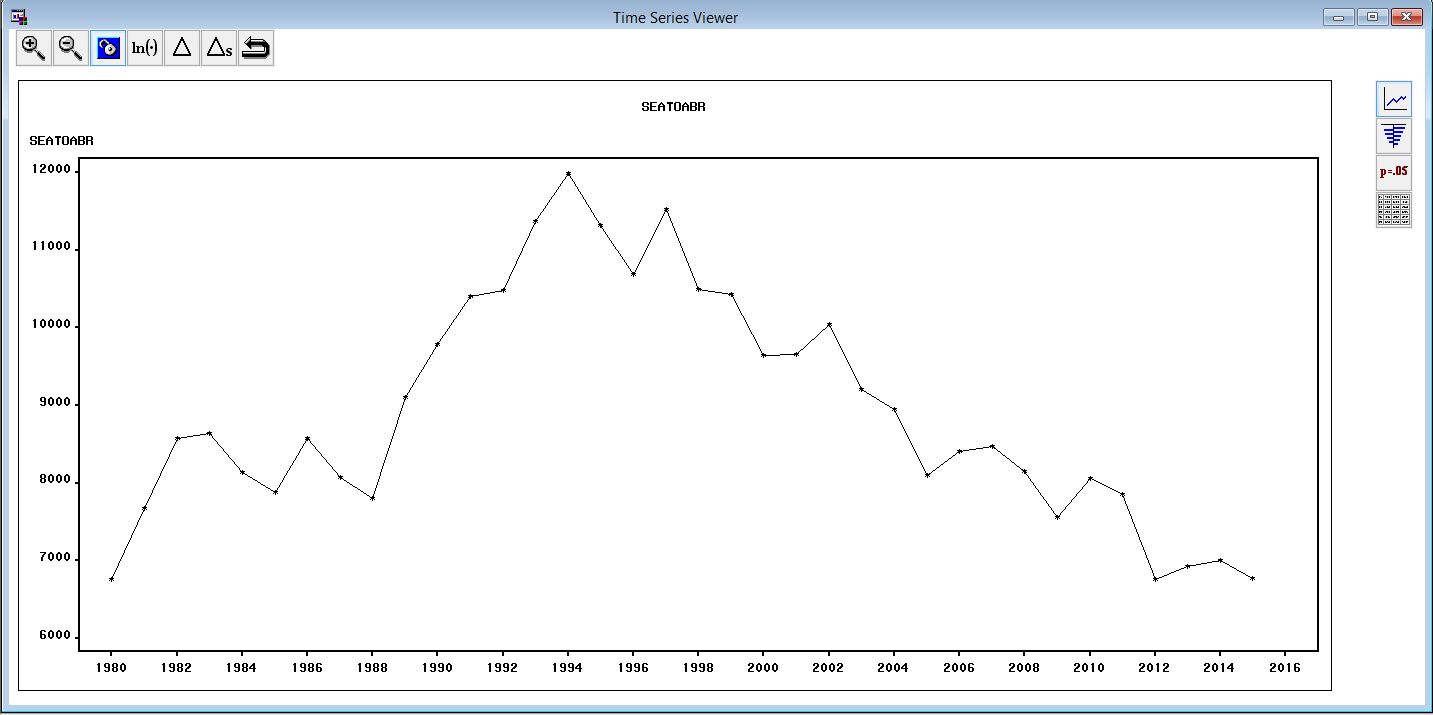


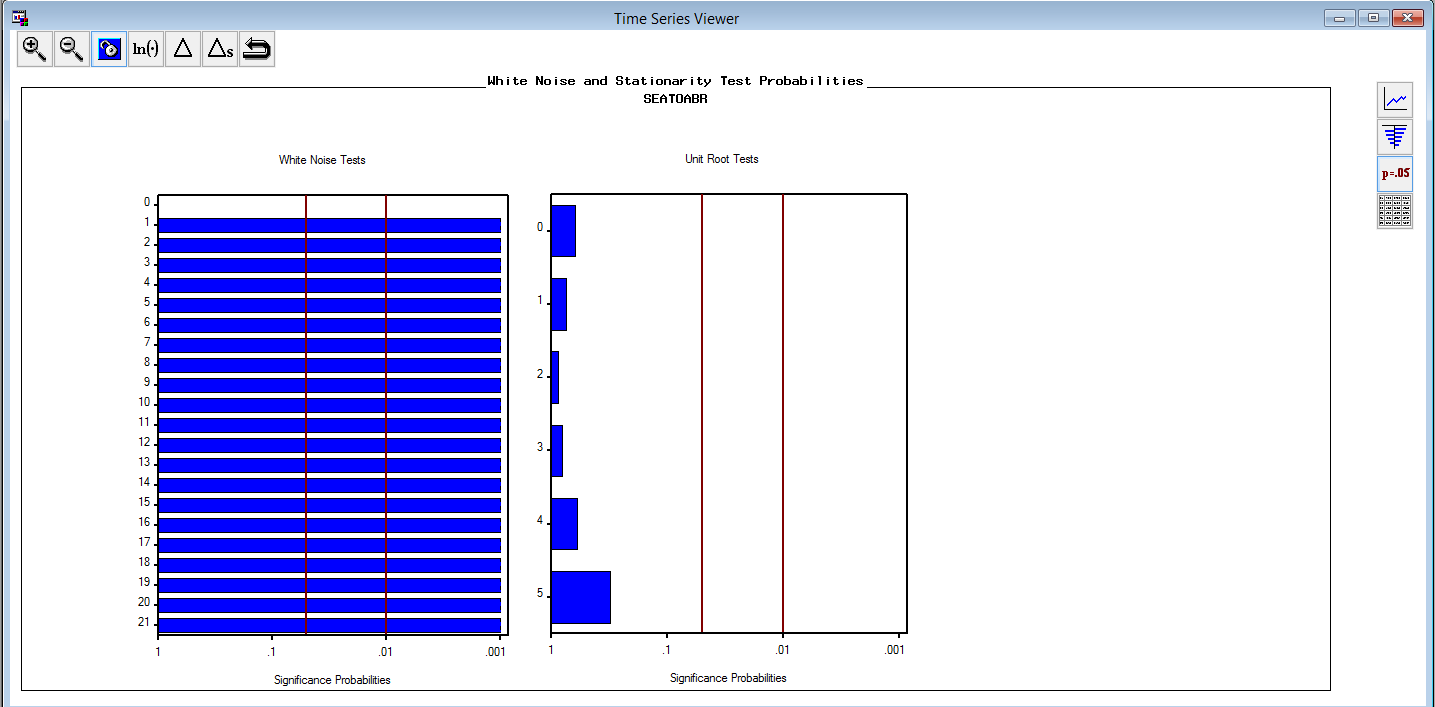




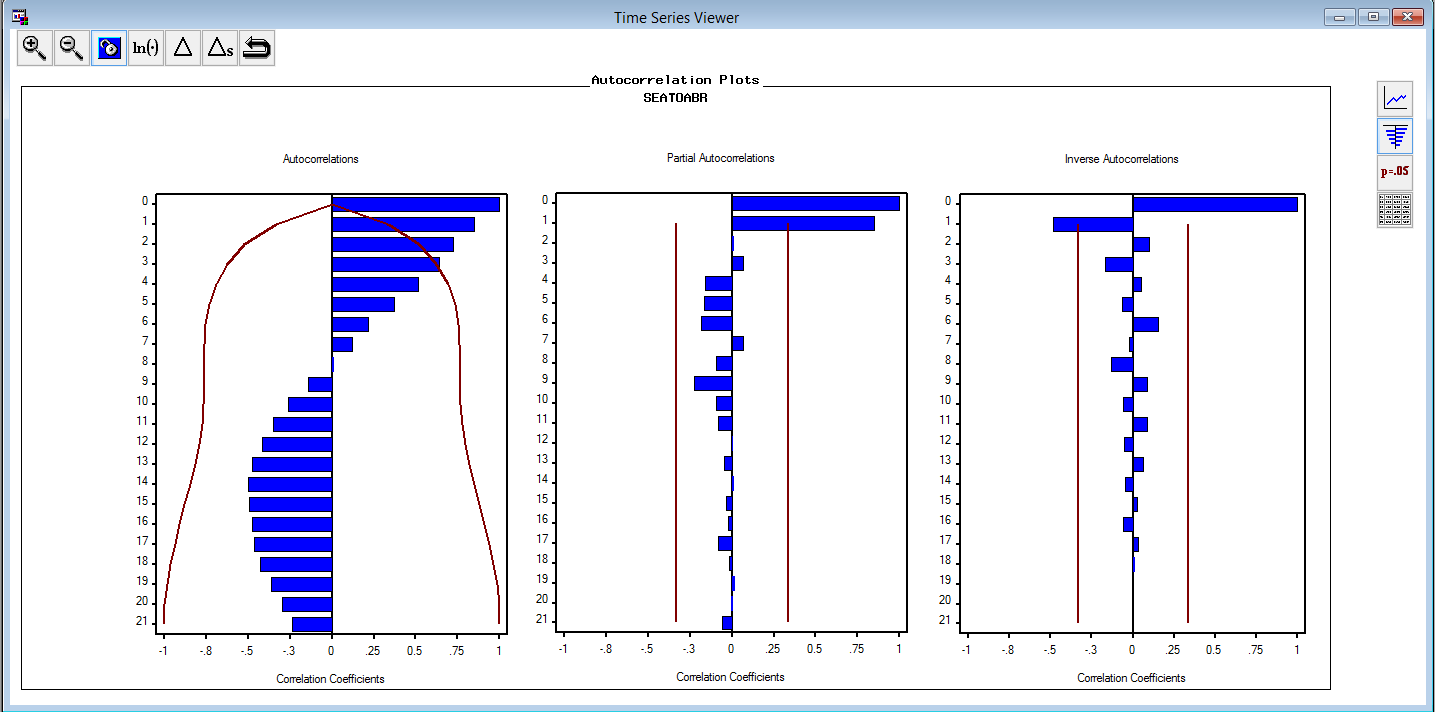


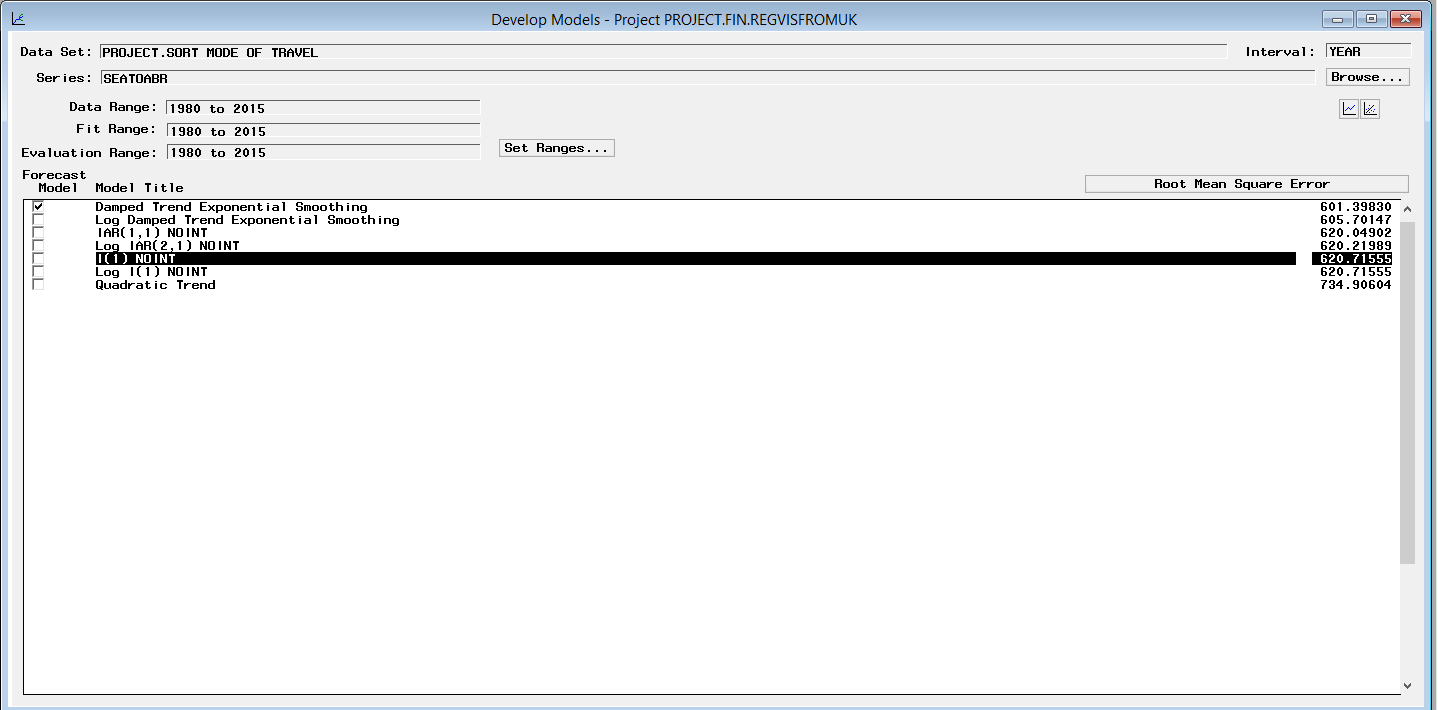
**UK Residents travel abroad by SEA**

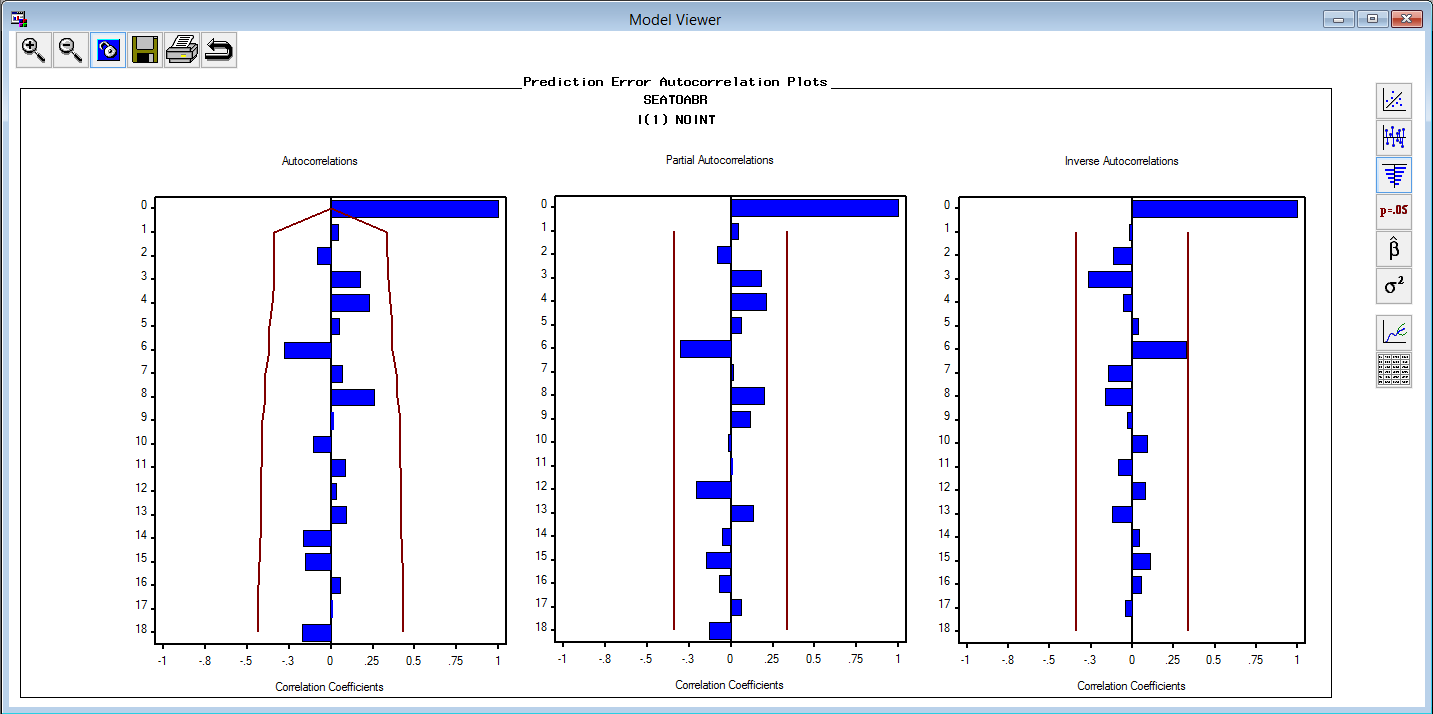


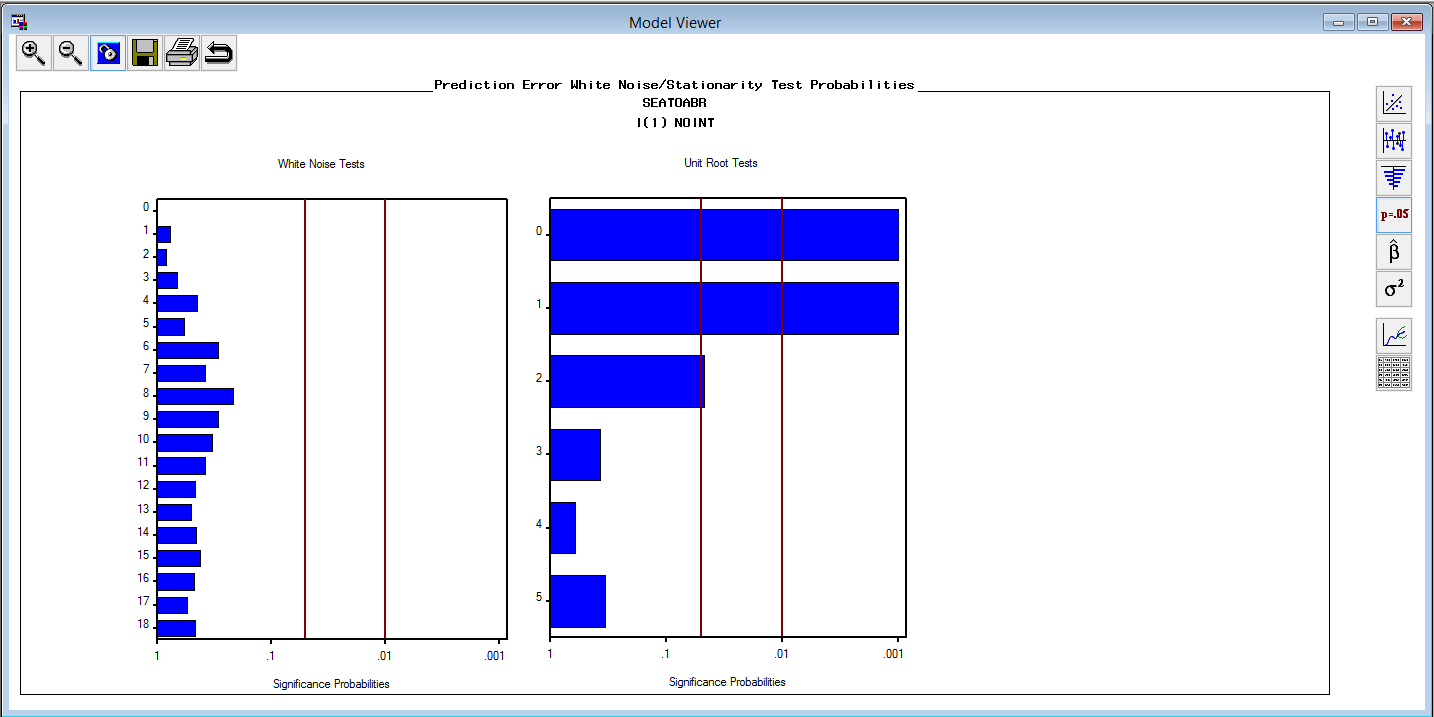


Many model combinations were tried, among which Log ARIMA(1,1,1) is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

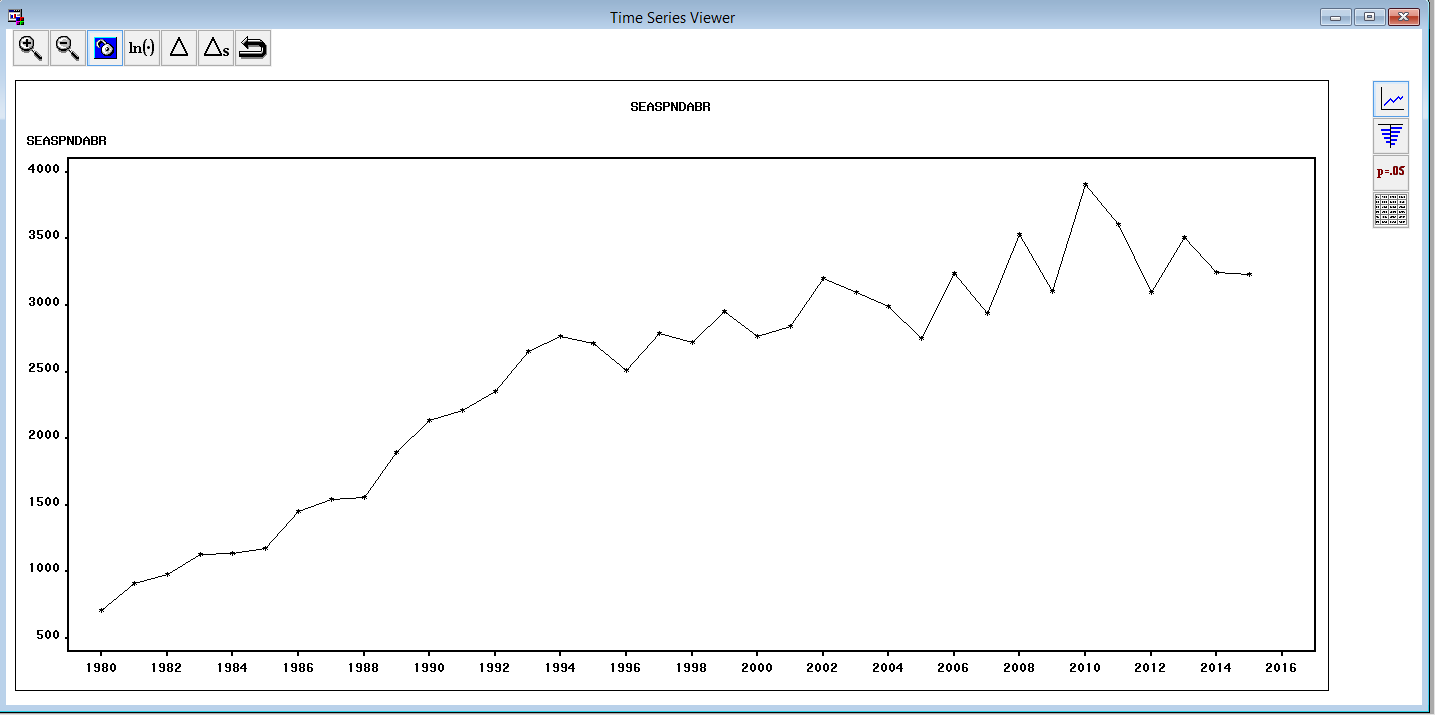
  
Many model combinations were tried, among which 1(1), built with just simple differencing is selected as the best fitting model after examining ACF, PACF and IACF plots etc.



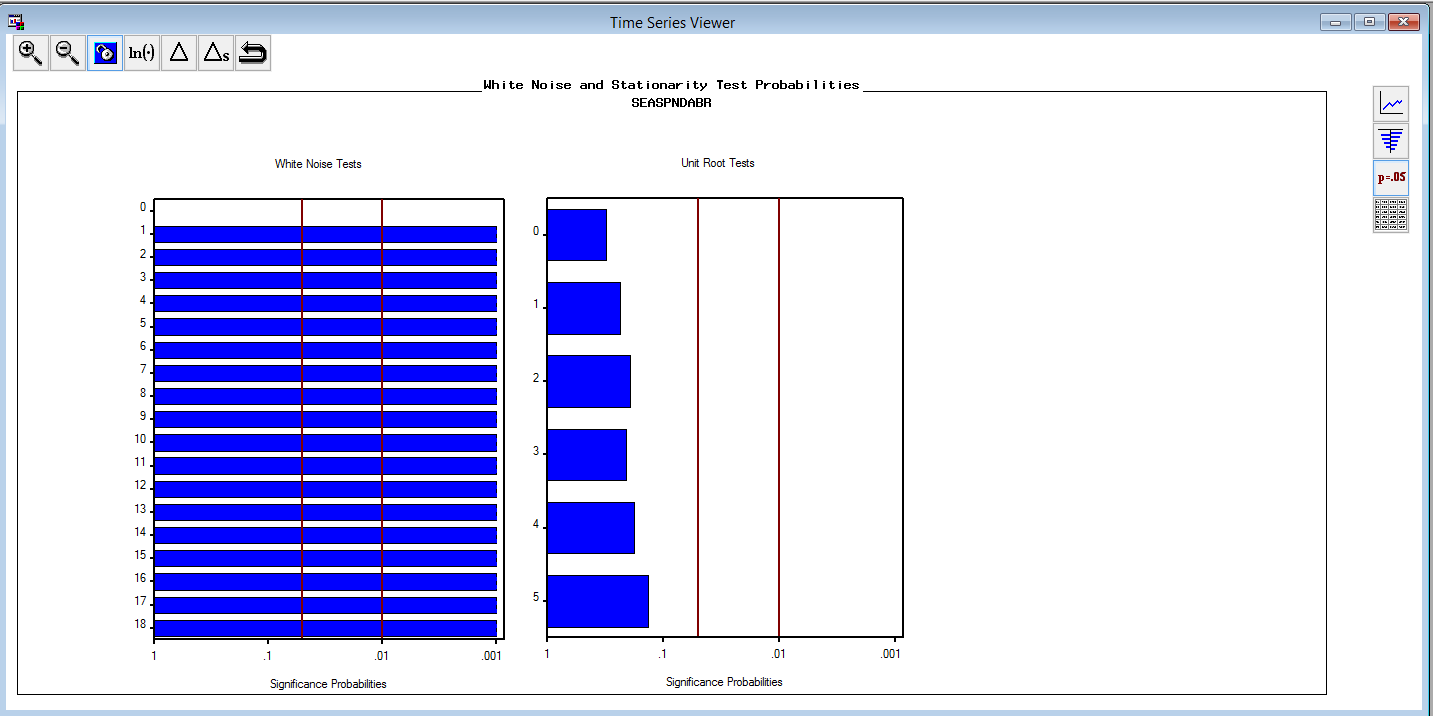




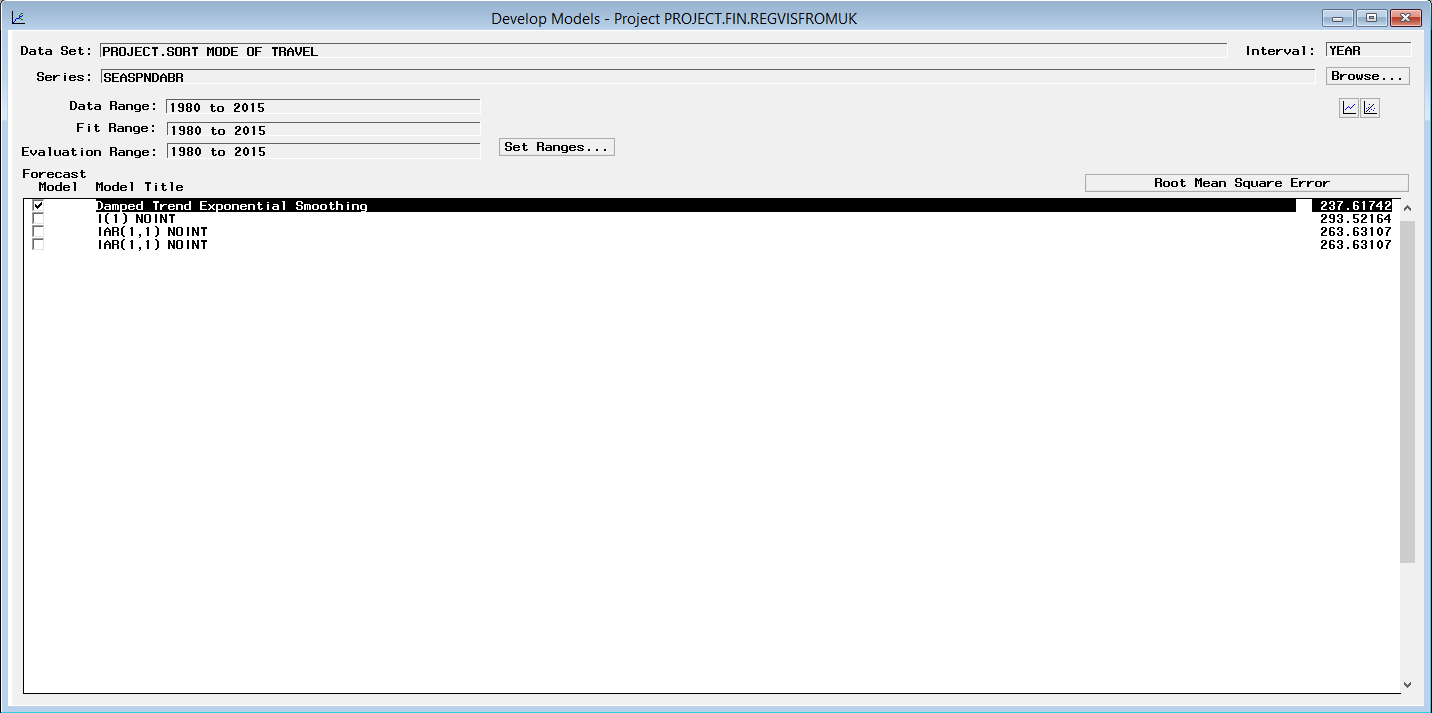
**UK Residents Spending on SEA Mode of Travel Abroad**

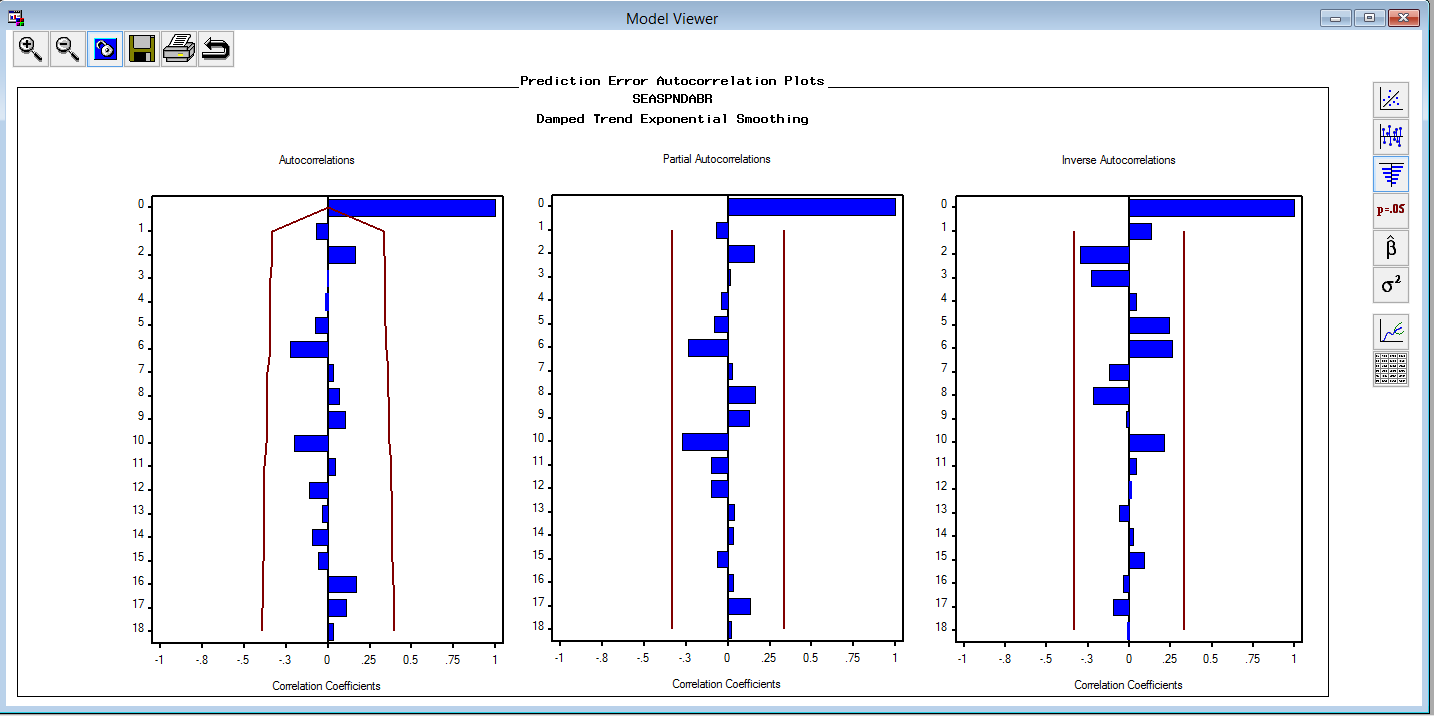


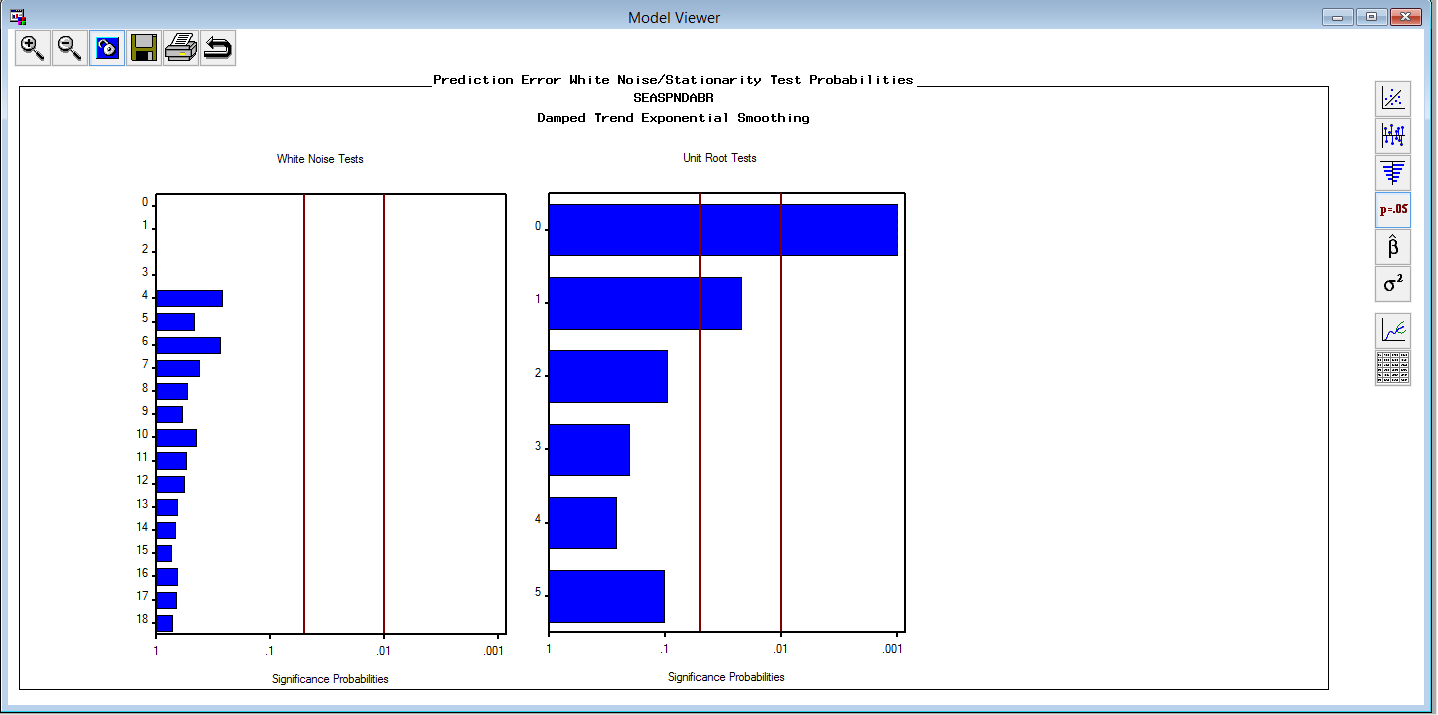




Many model combinations were tried, among which Damped Trend Exponential Smoothing is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

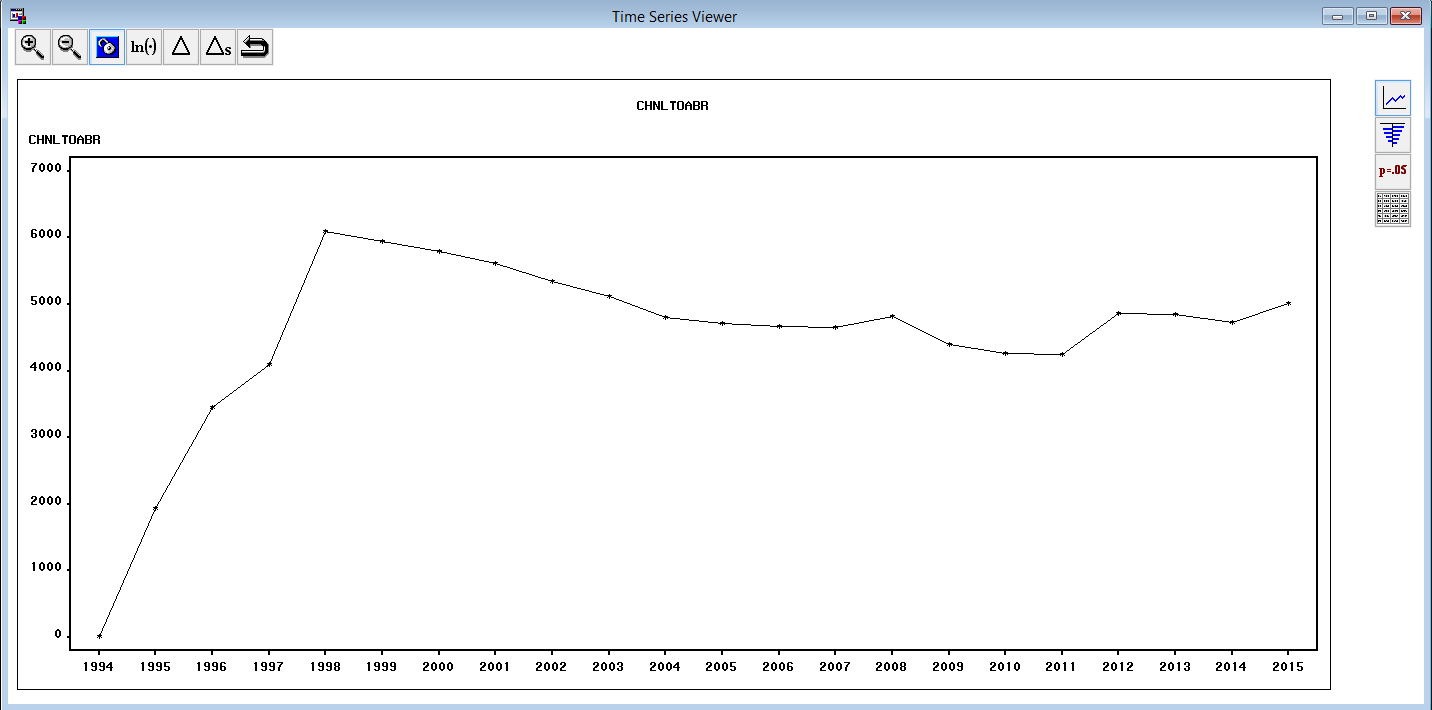


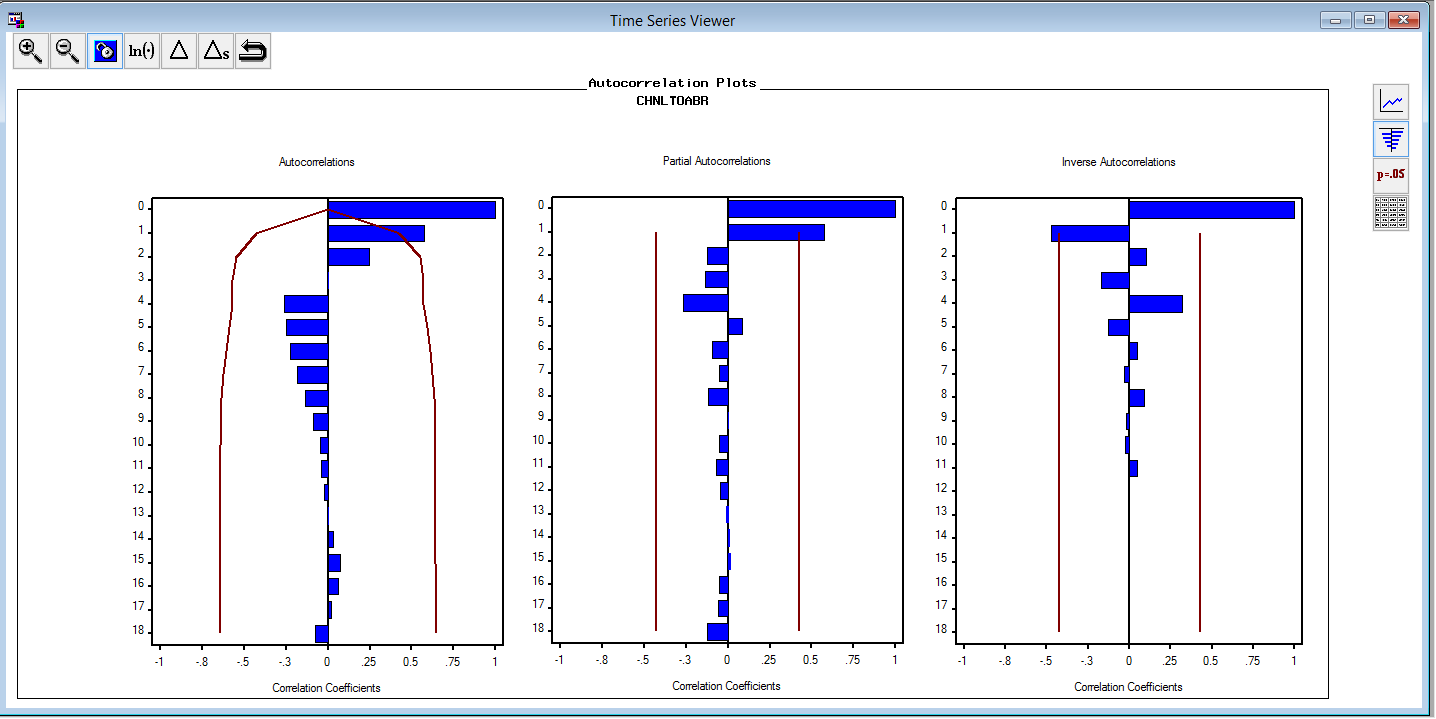


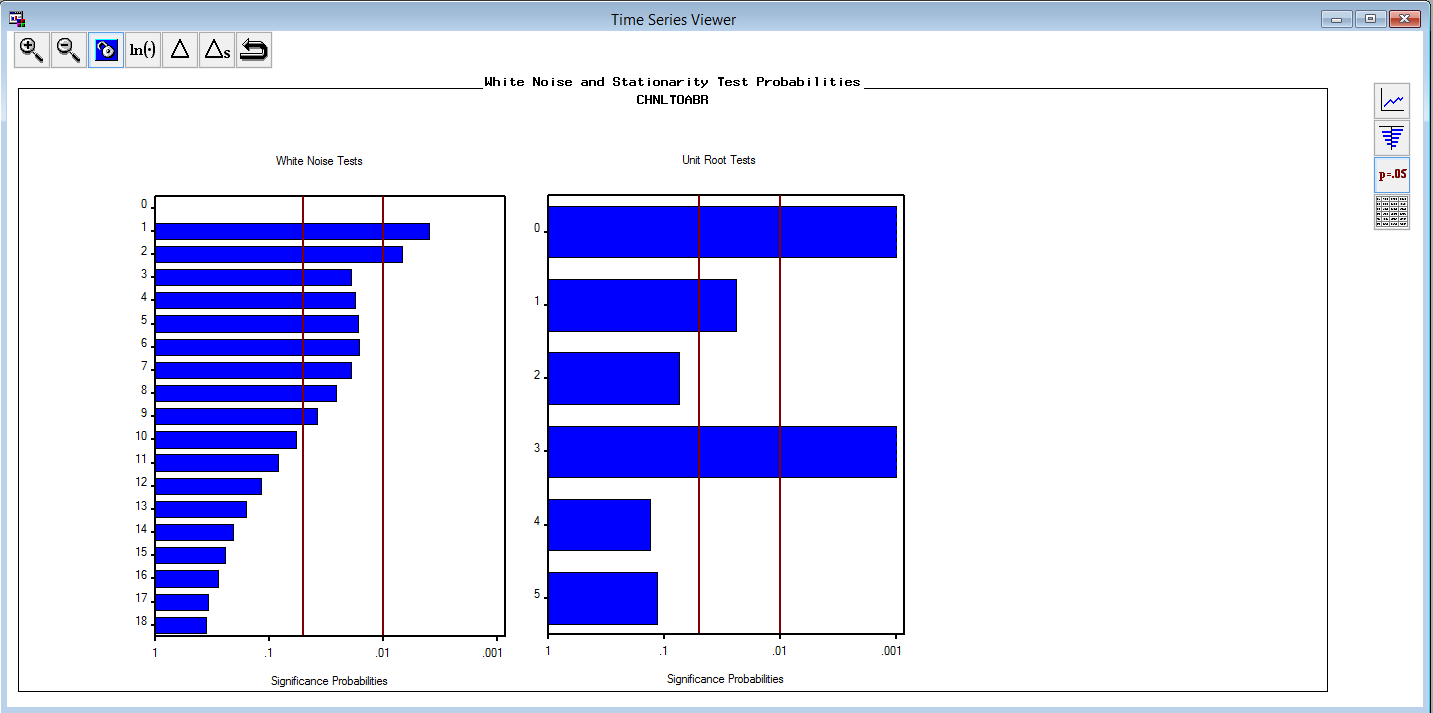




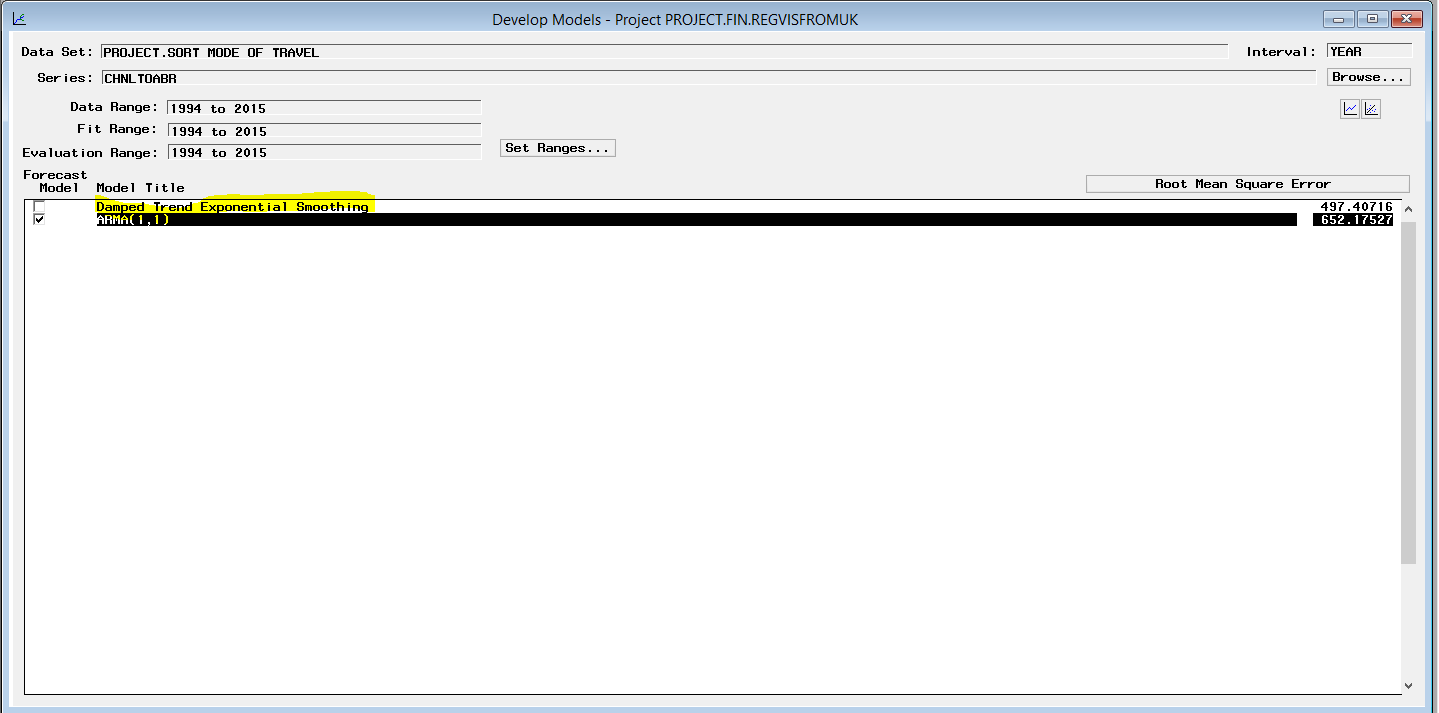
**UK Residents travel abroad through Channel**

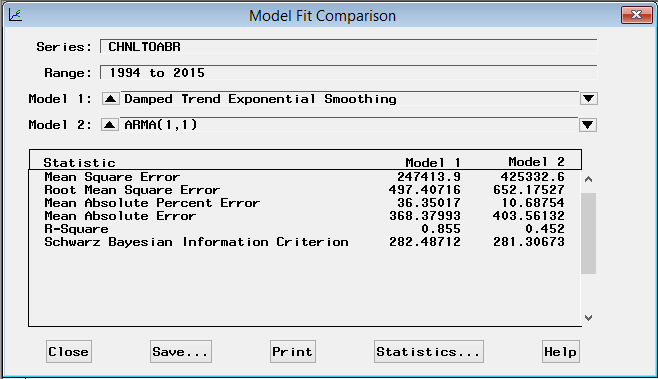


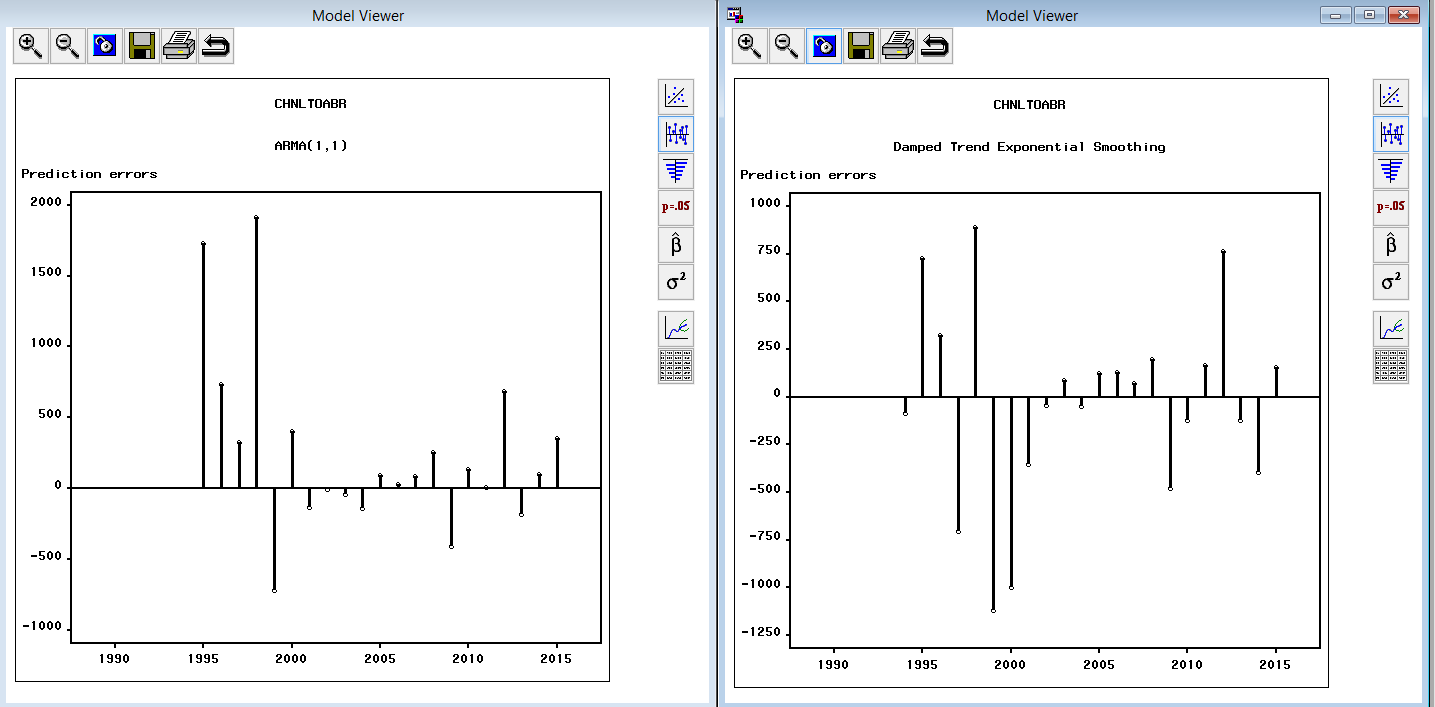


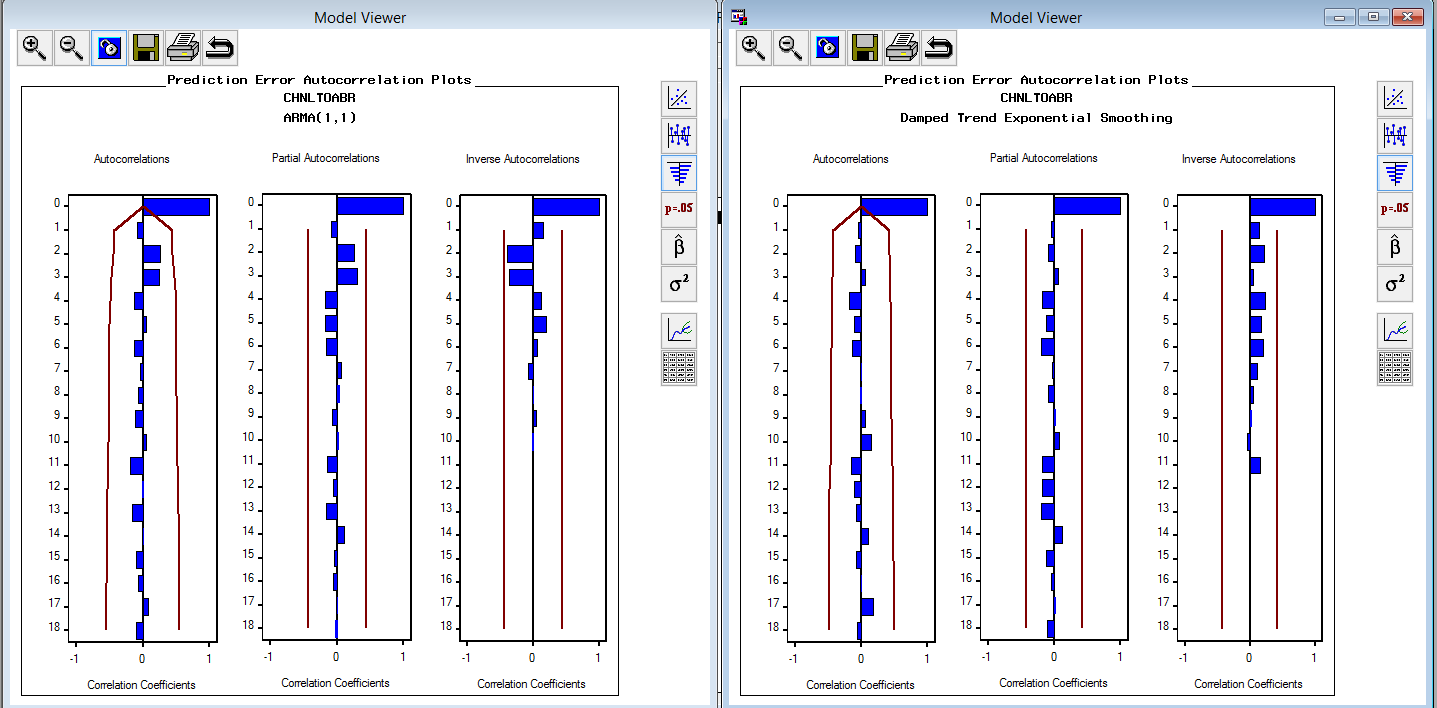


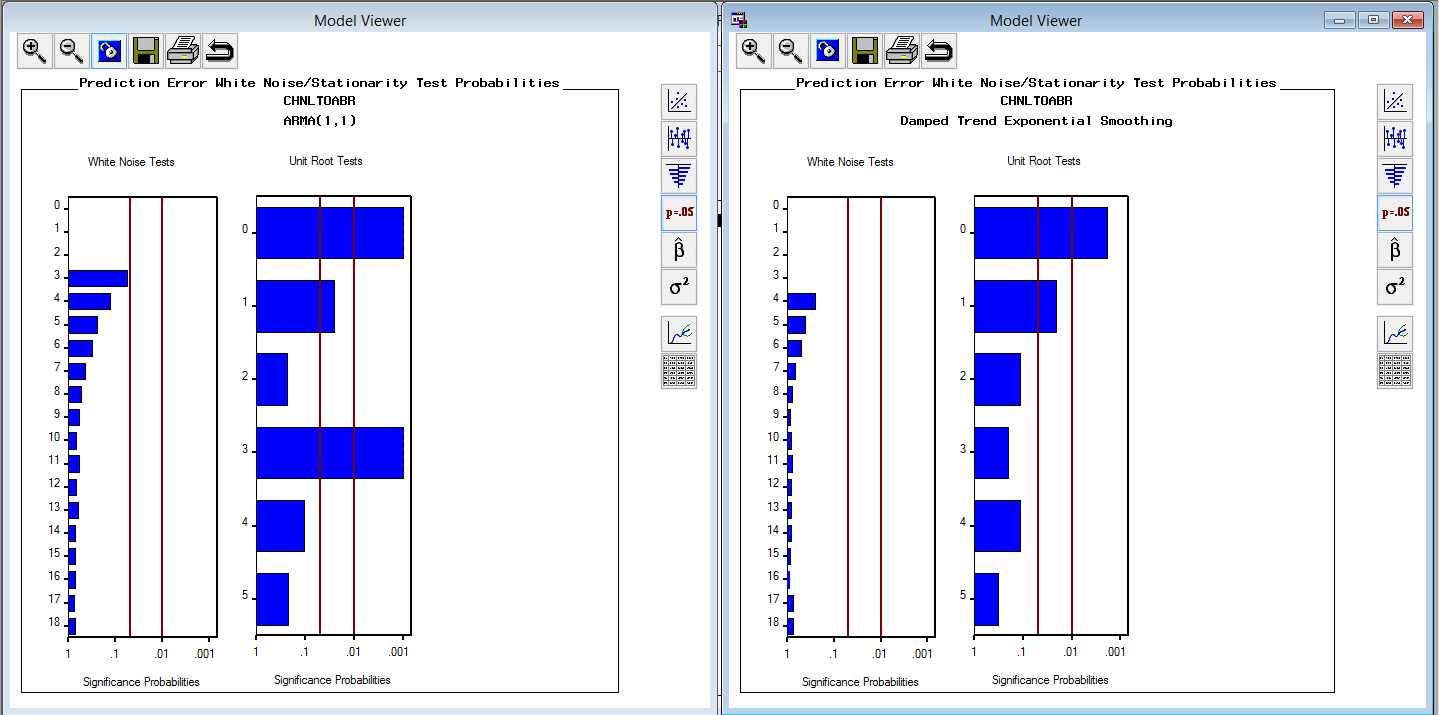
Since, the model is already stationary, hence, no first differencing required. Many model combinations were tried, among which ARMA (1,1) is selected as the best fitting model after examining ACF, PACF and IACF plots etc.



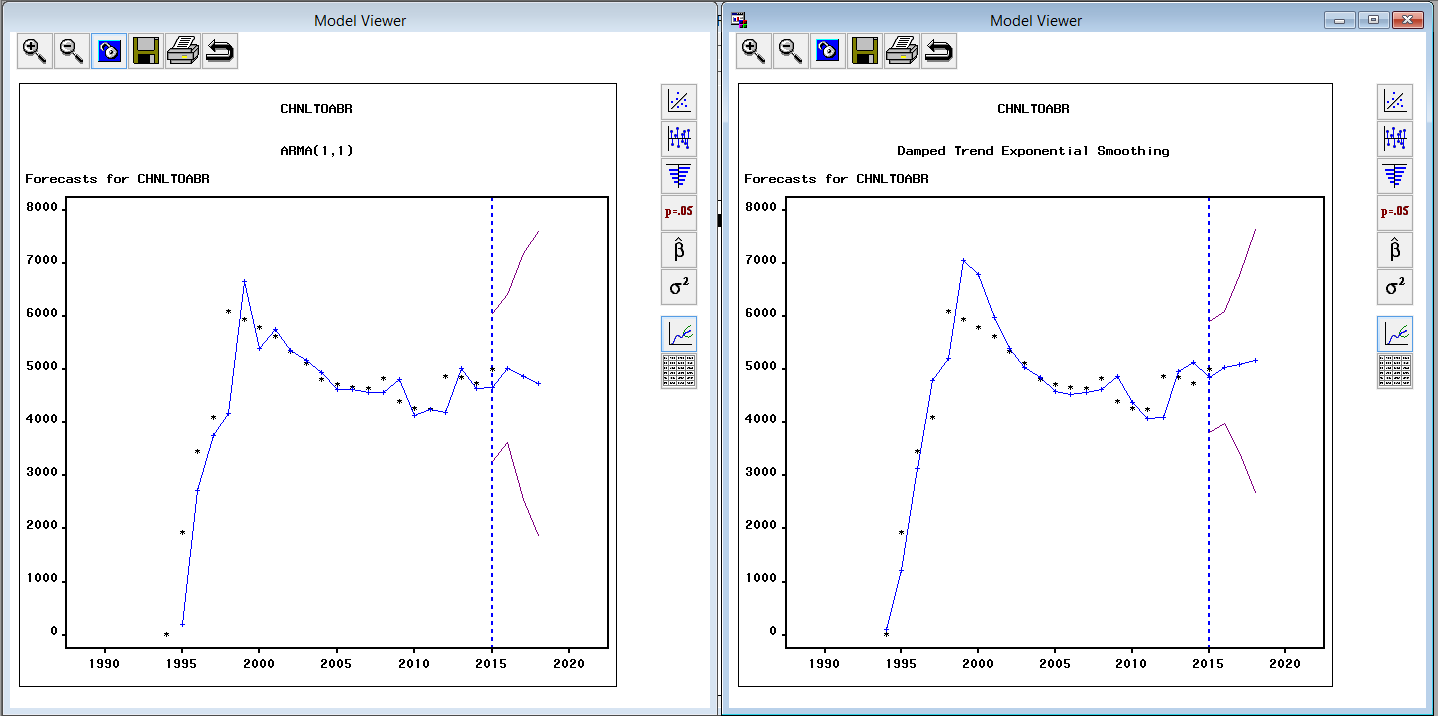




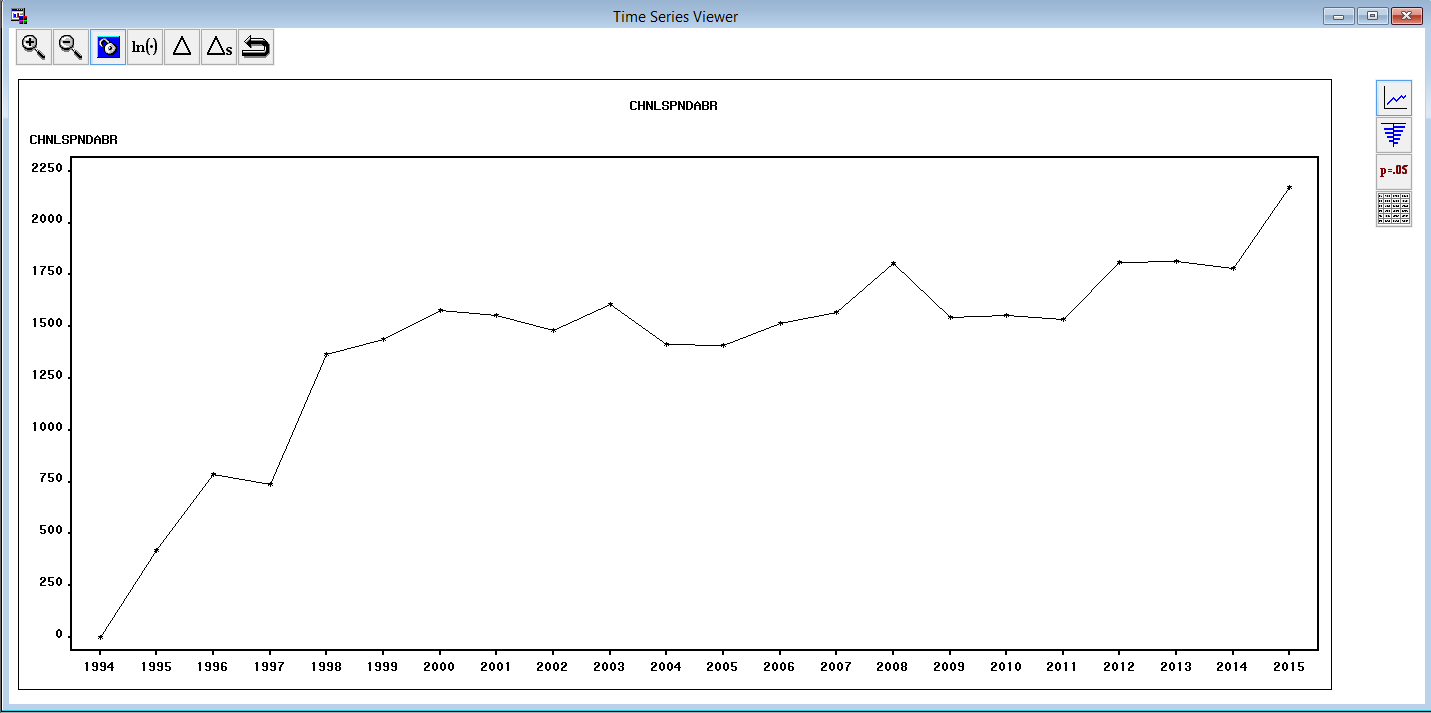


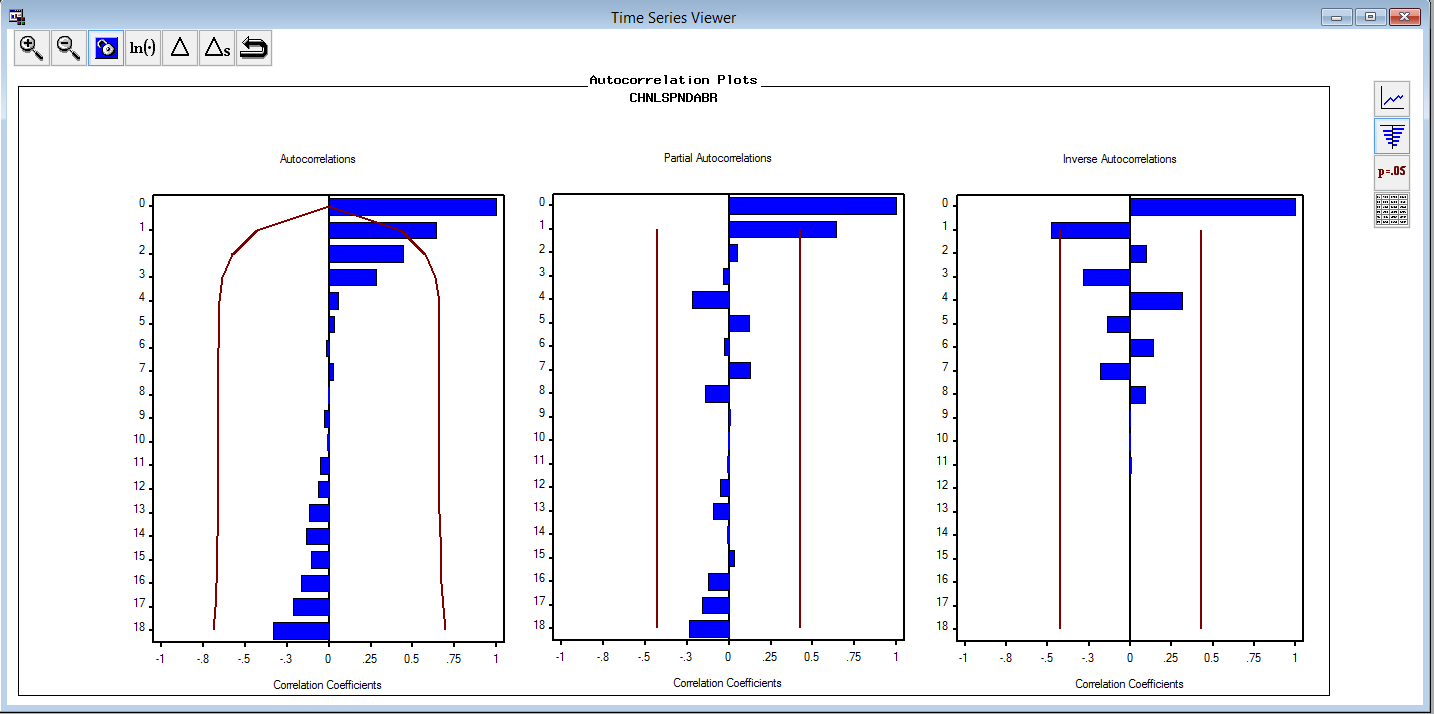


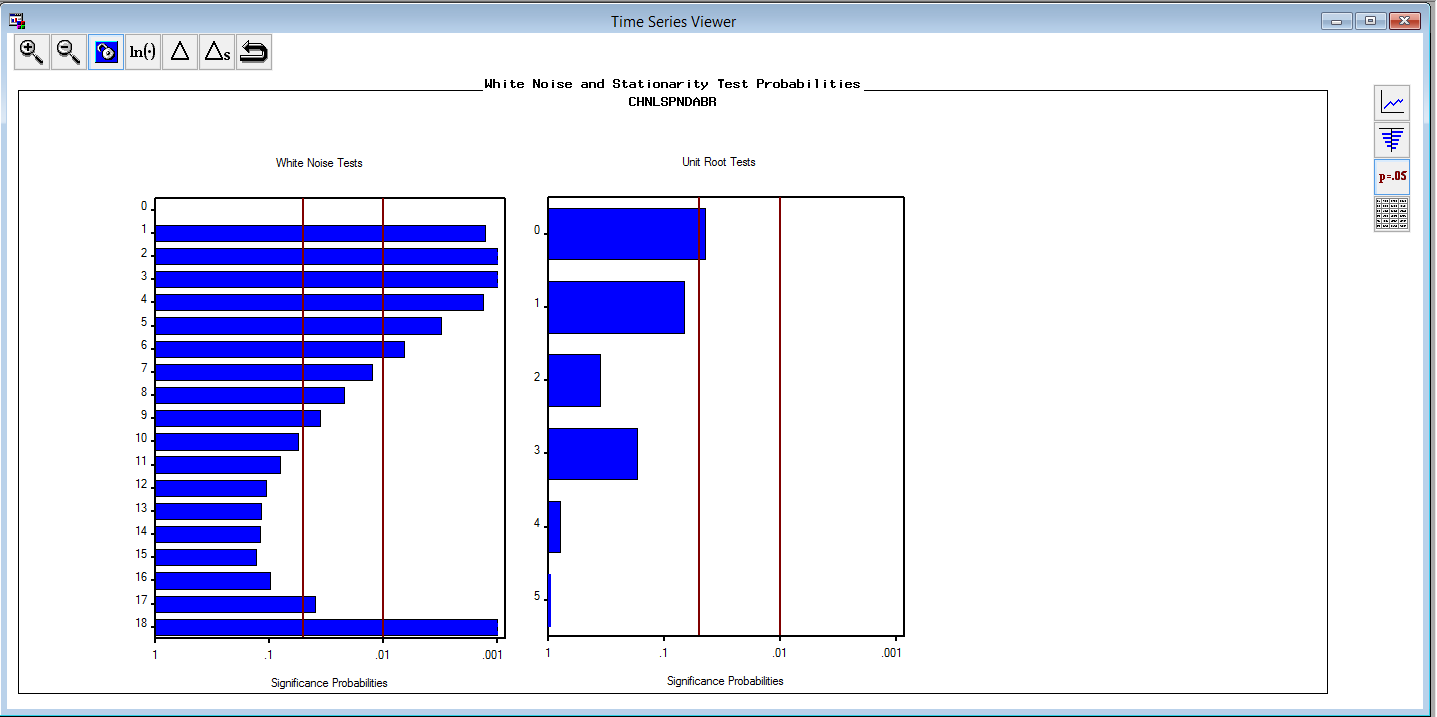
Damped exponential smoothing performs poor in stationarity test.



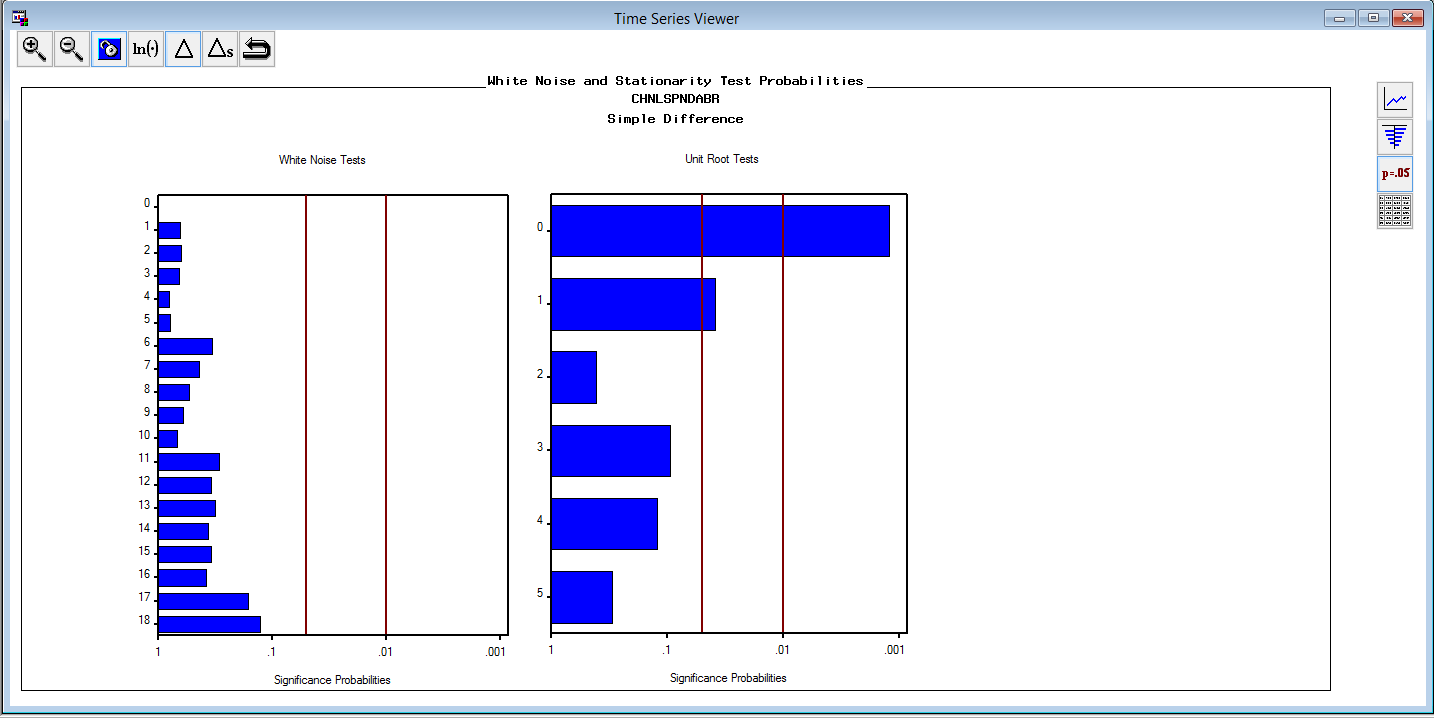
**UK Residents Spending on Channel Mode of Travel Abroad**

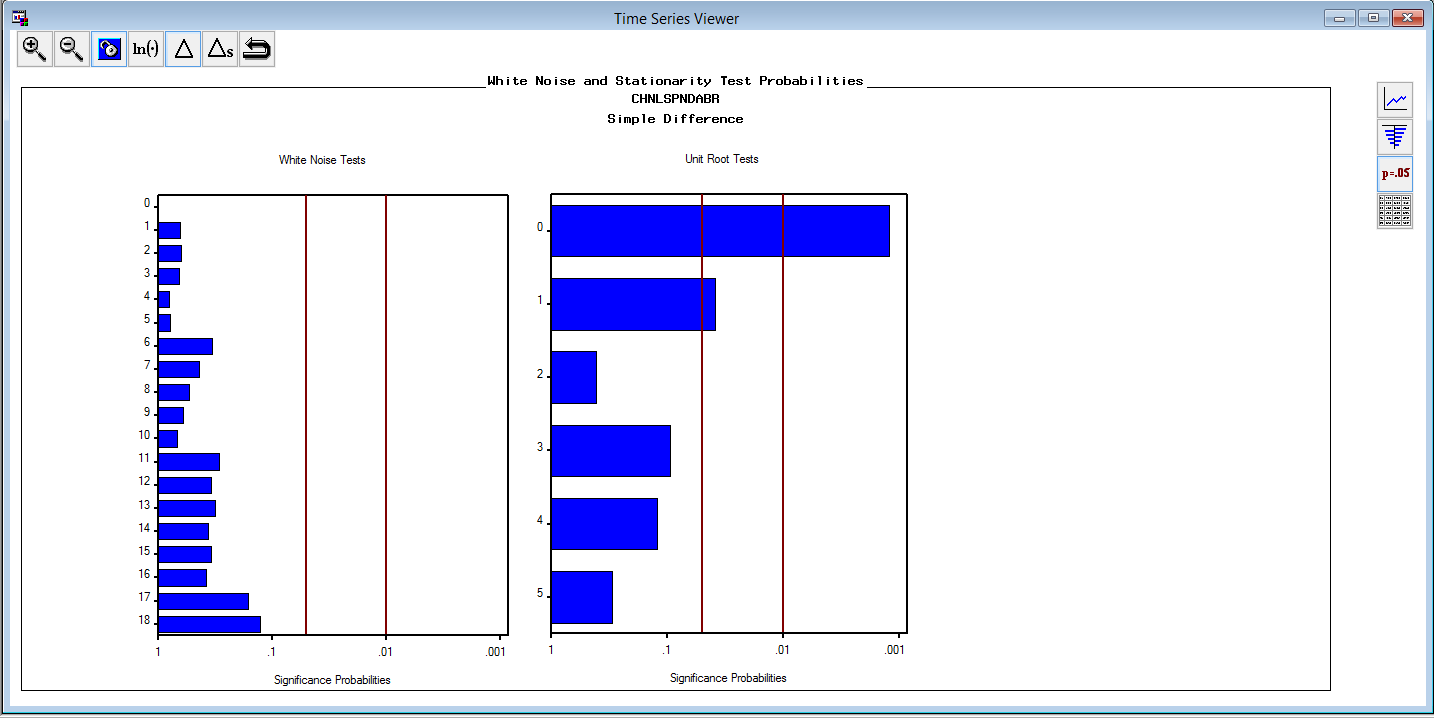




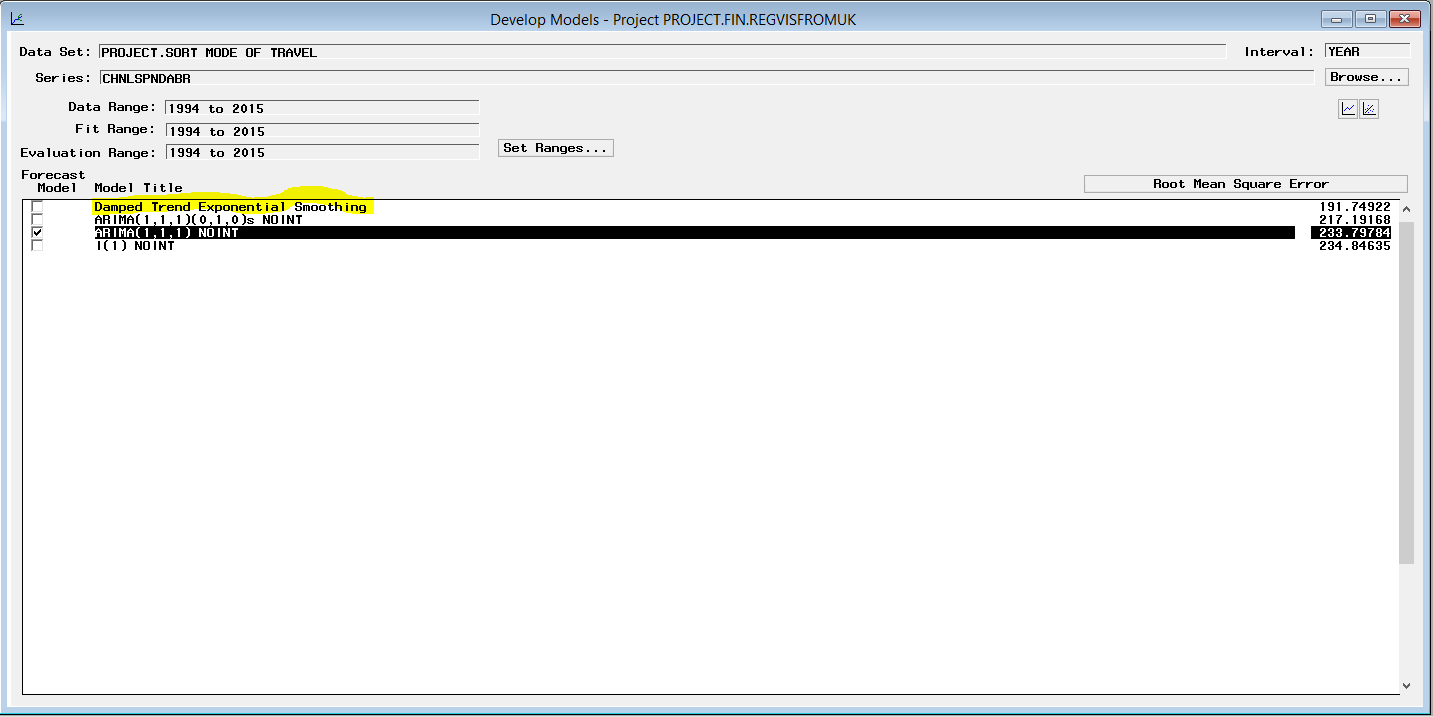


After first differencing;

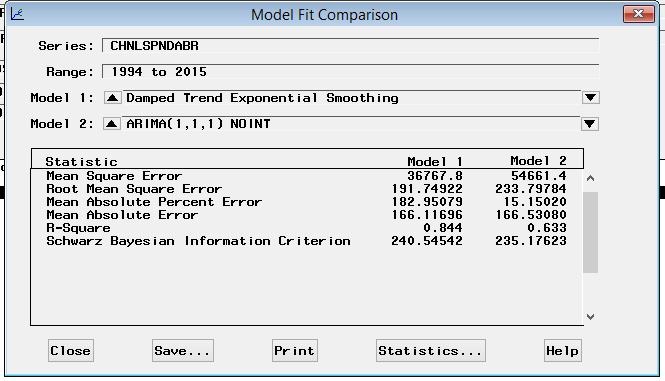


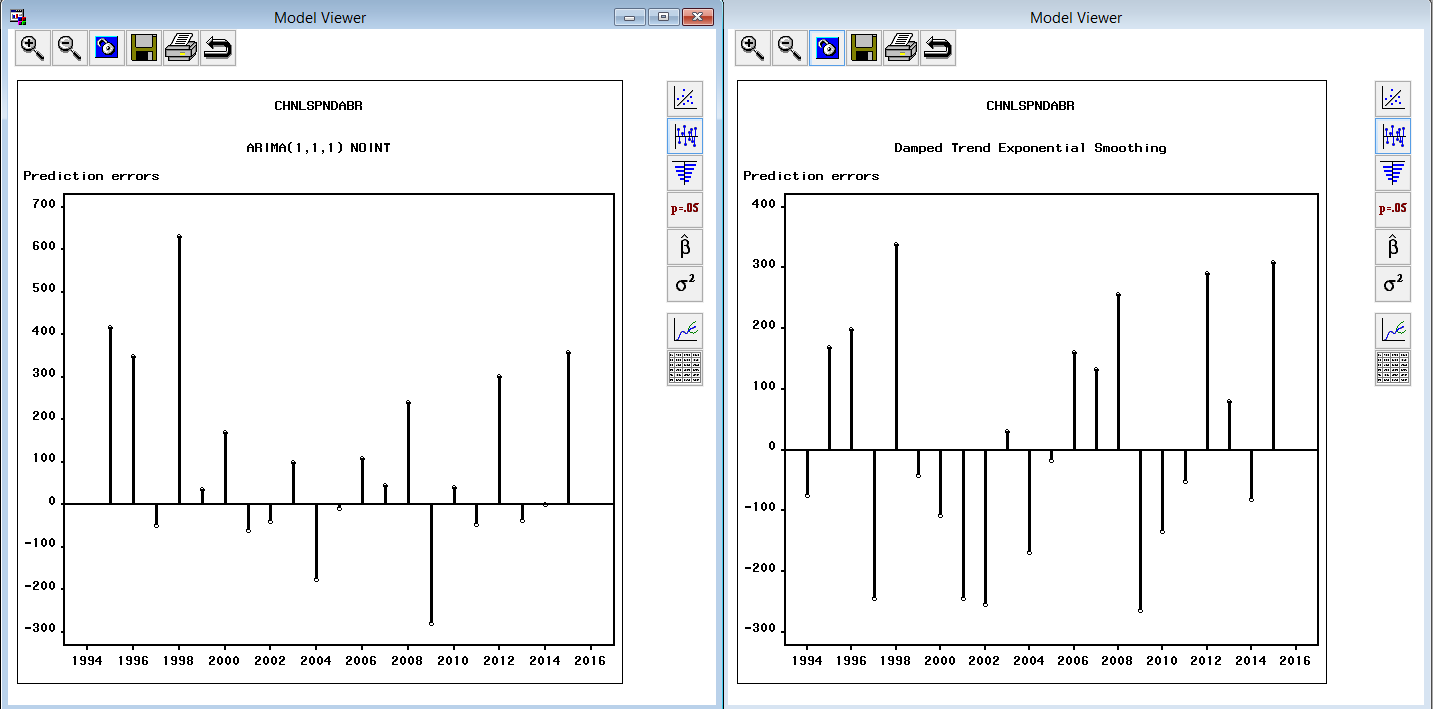


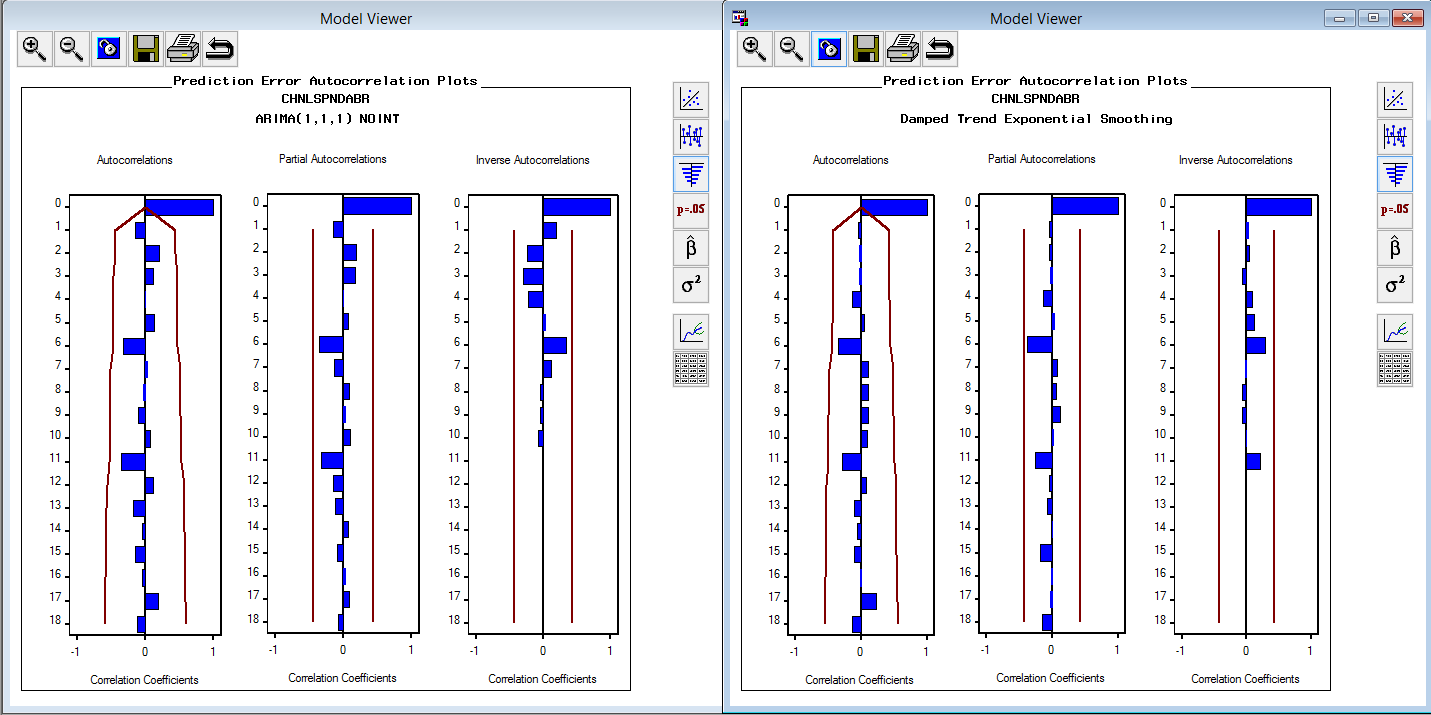
Many model combinations were tried, among which ARIMA ( 1,1,1) is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

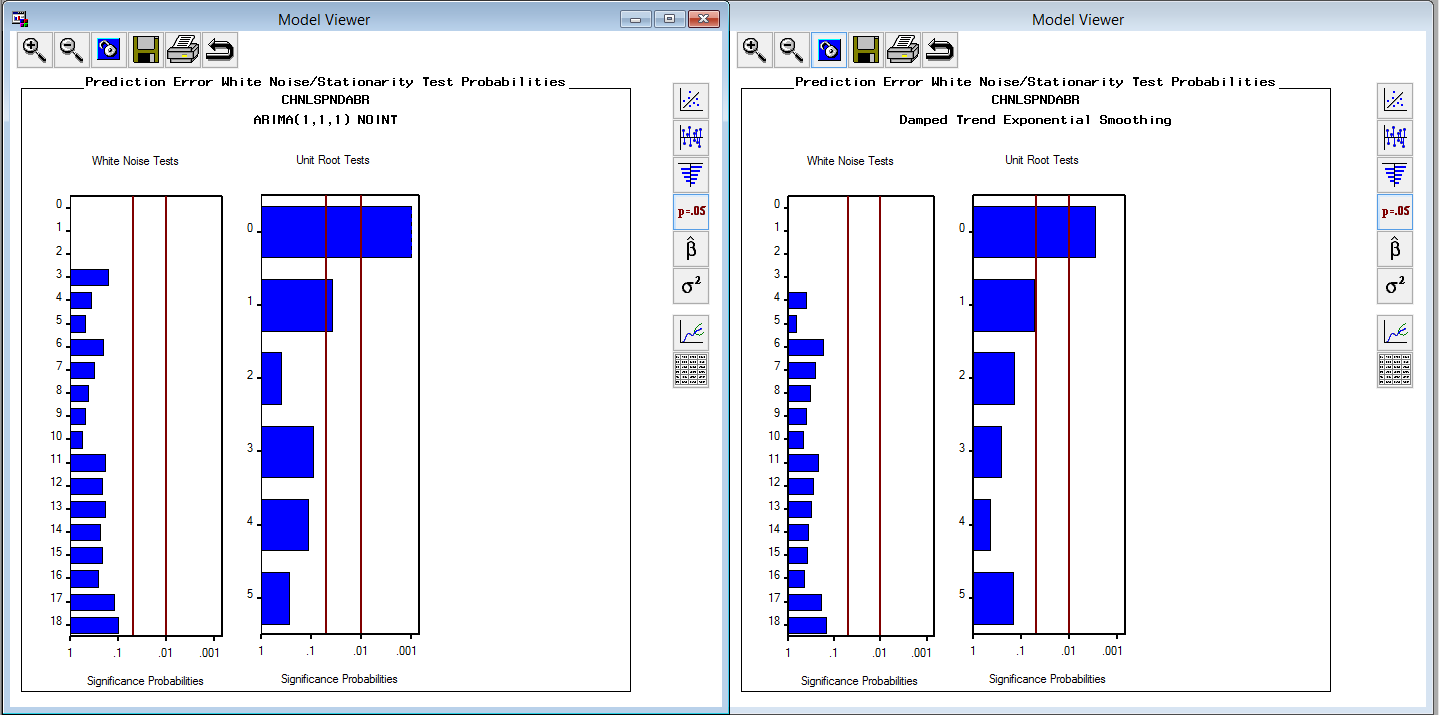


Comparison between two models







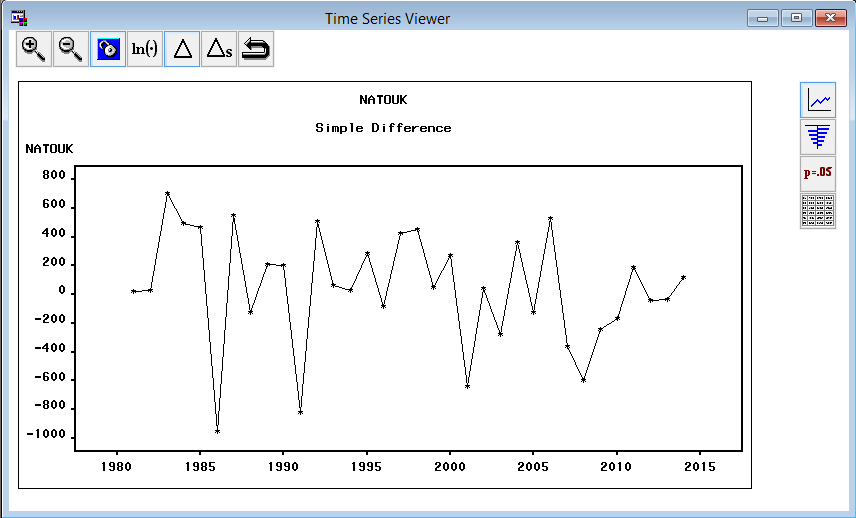




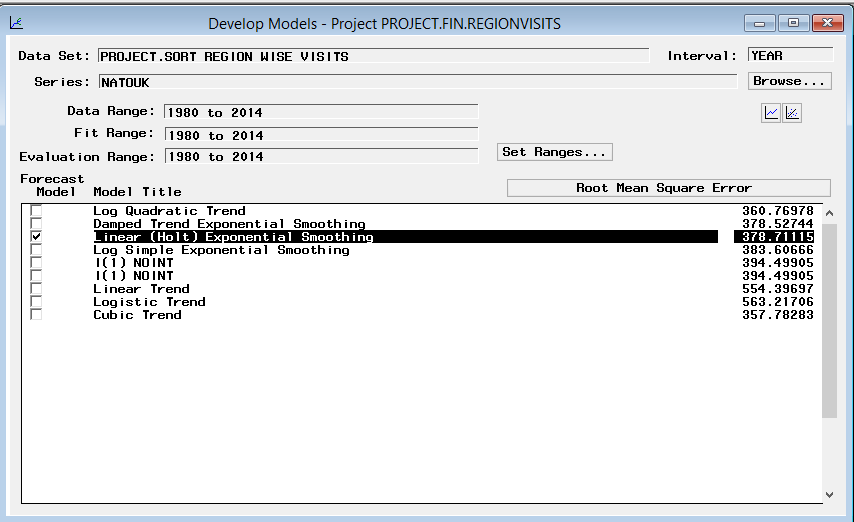
Hence, we select ARIMA(1,1,1) as damped exponential smoothing performs poor in stationarity test

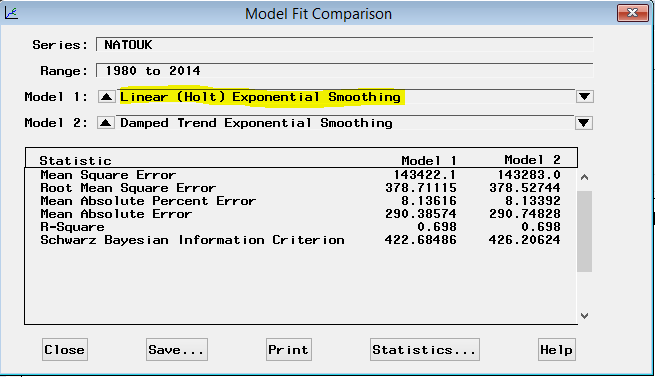
**Overseas residents visit to UK**

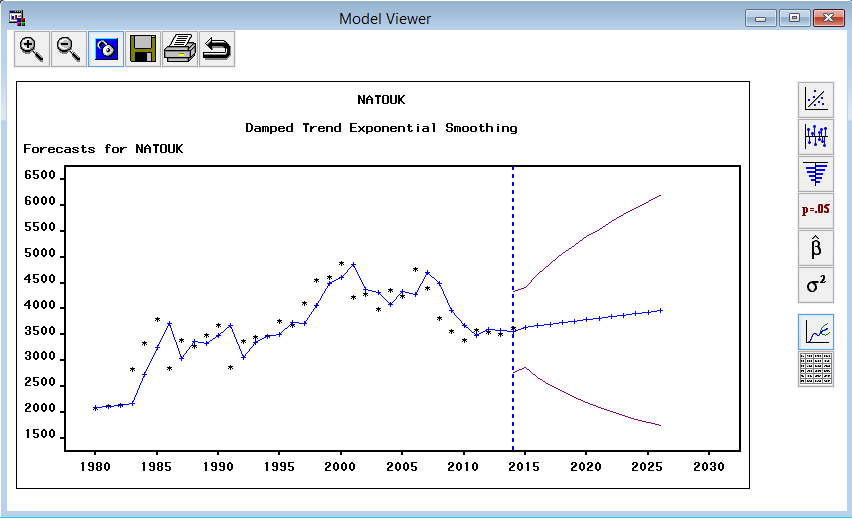
Residents from North America visiting UK



Many model combinations were tried, among which Linear(HOLT) Exponential Smoothing is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

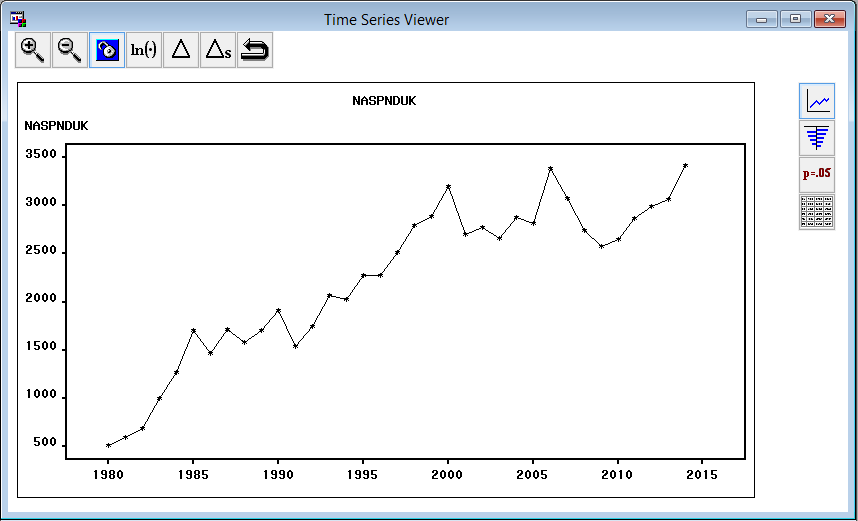




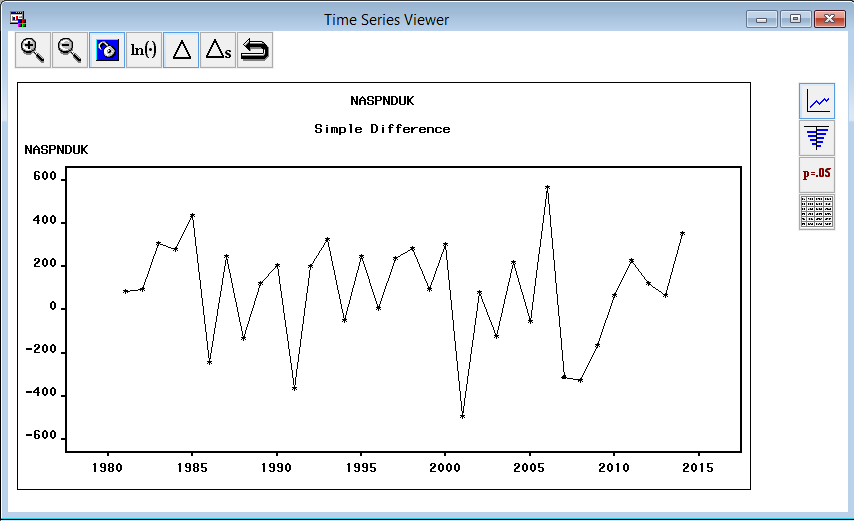


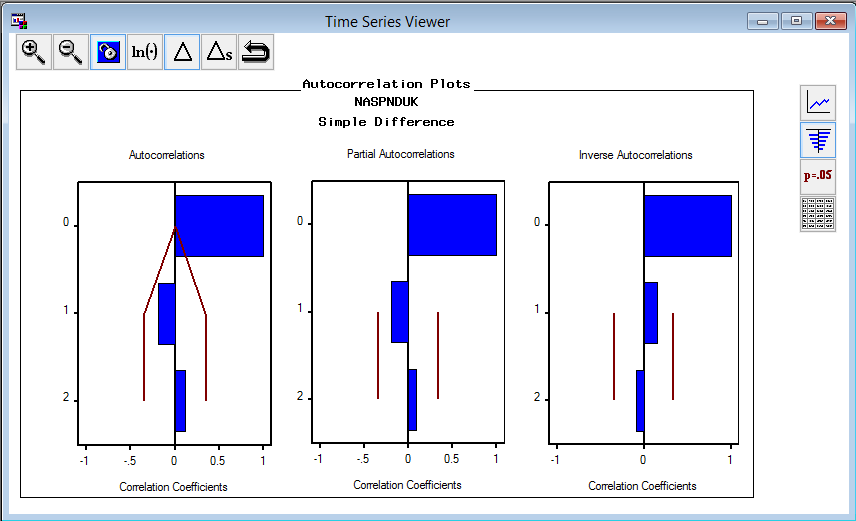
We don’t need the model with large confidence interval. Hence, we reject this model.

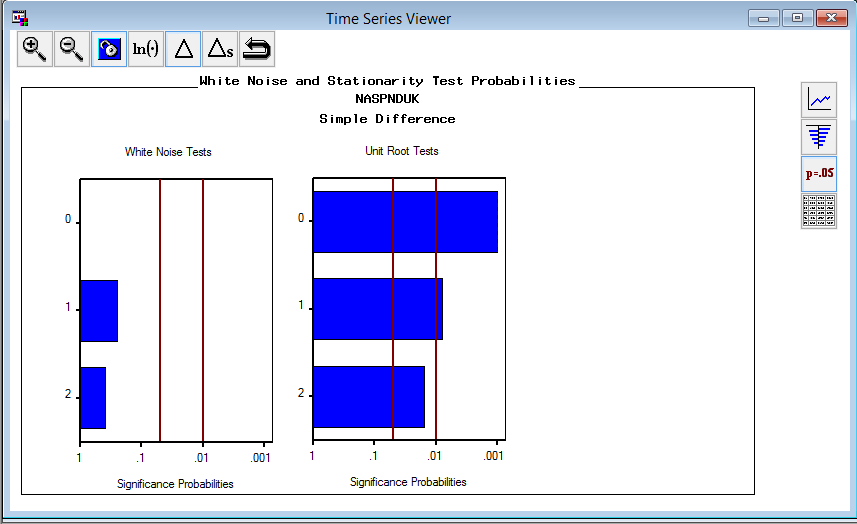
**Residents from North America spending in UK**



After first differencing:

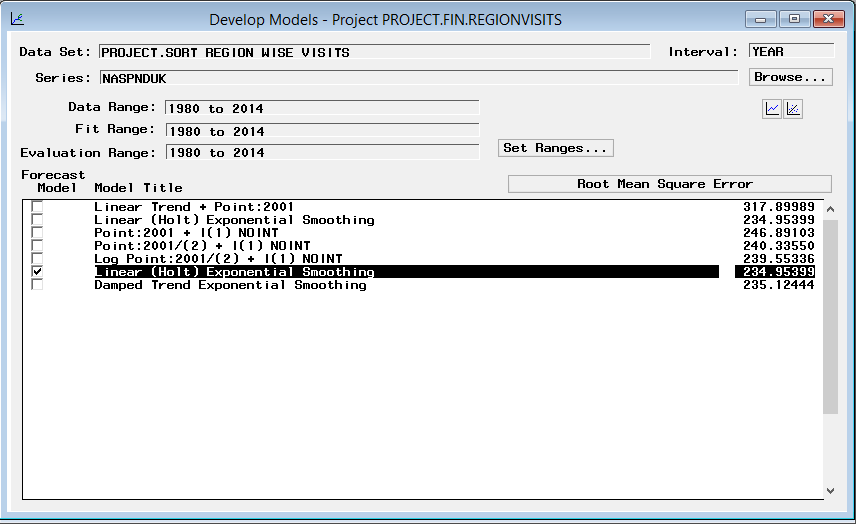


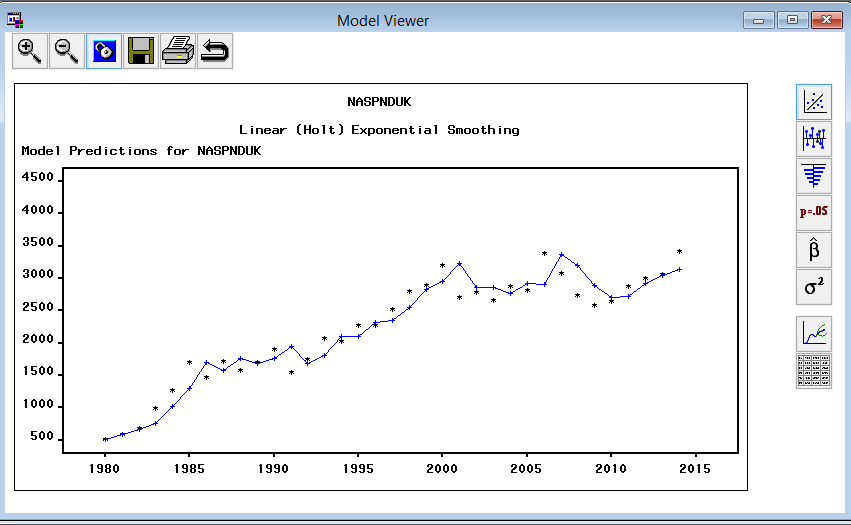




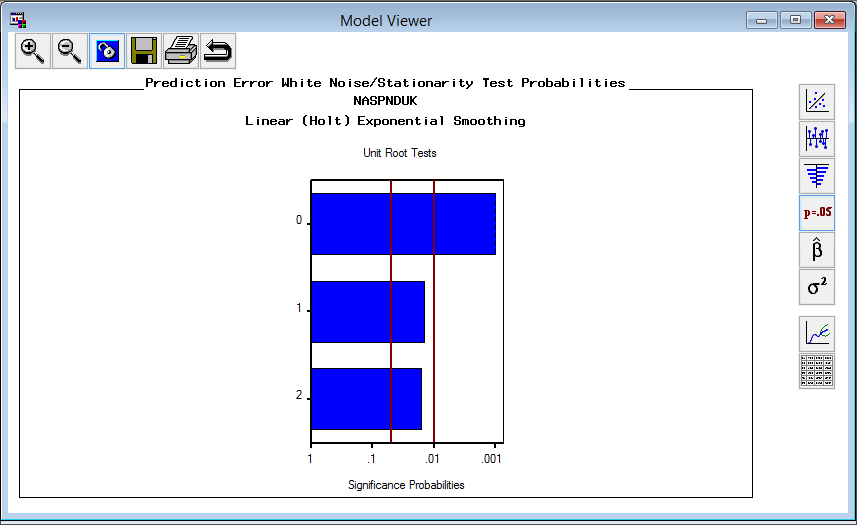
Hence, we will apply first differencing while we build our model.

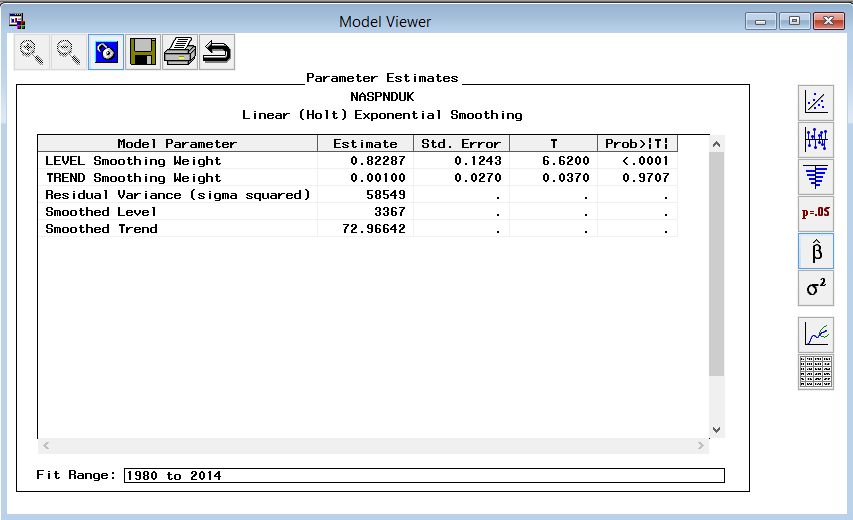
Many model combinations were tried, among which Linear (Holt) Exponential is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

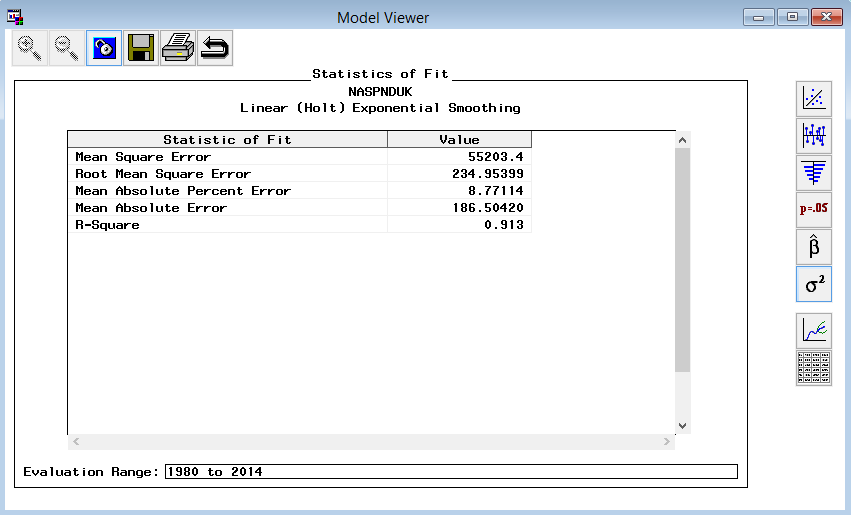


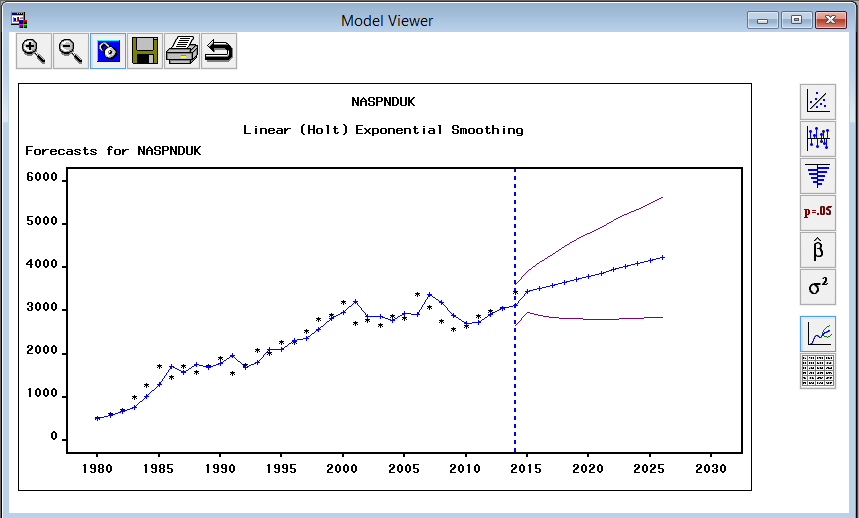


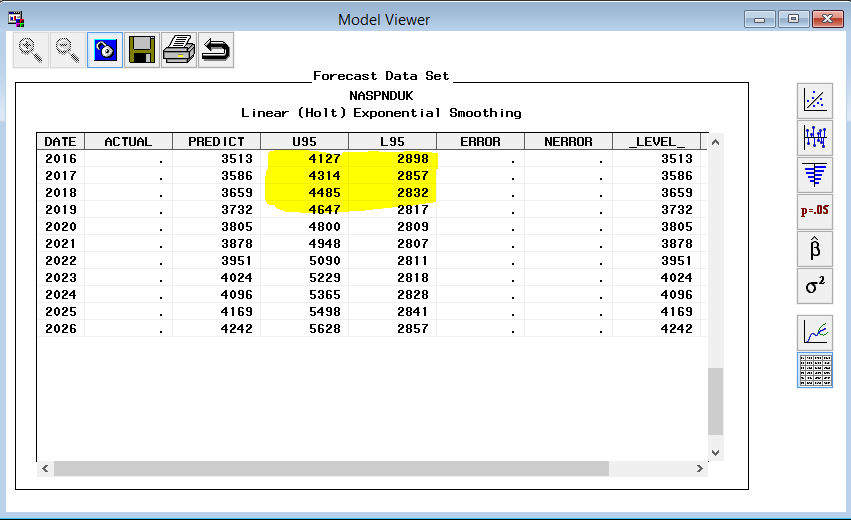




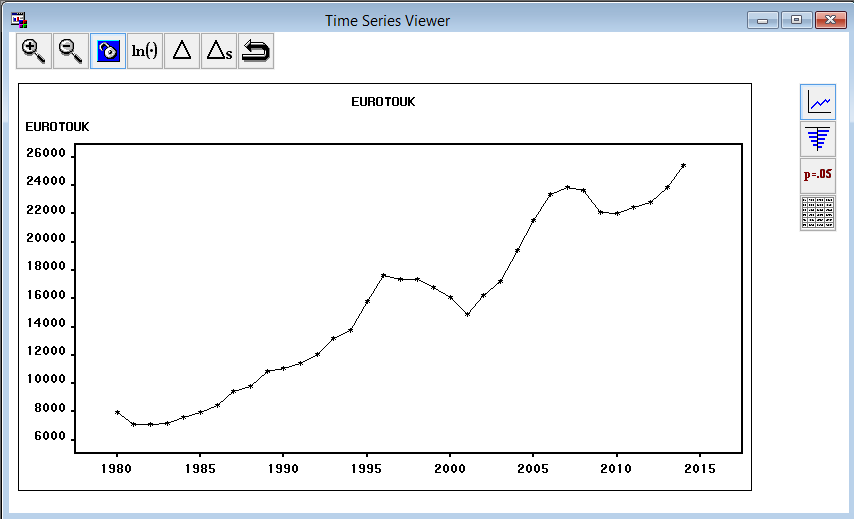




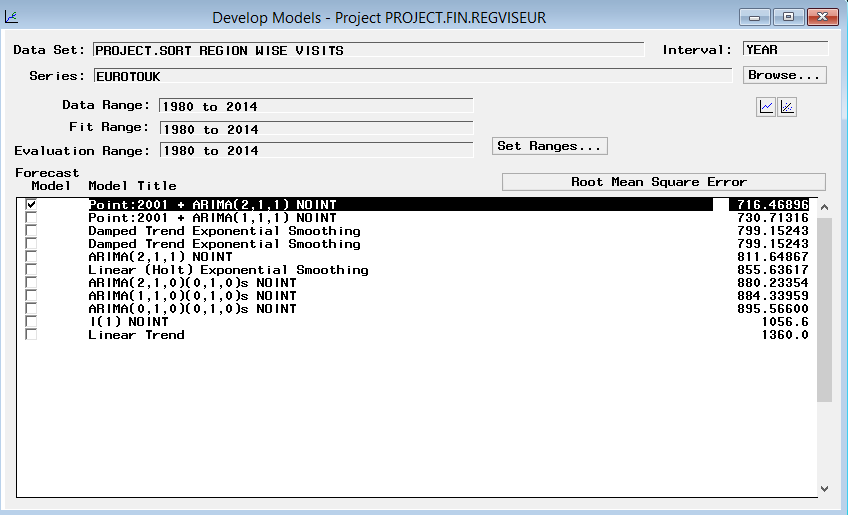


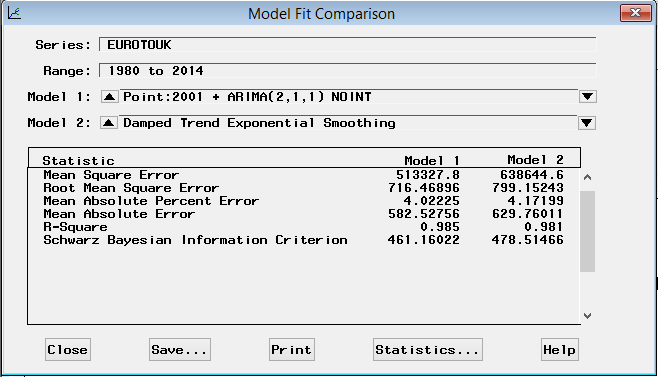


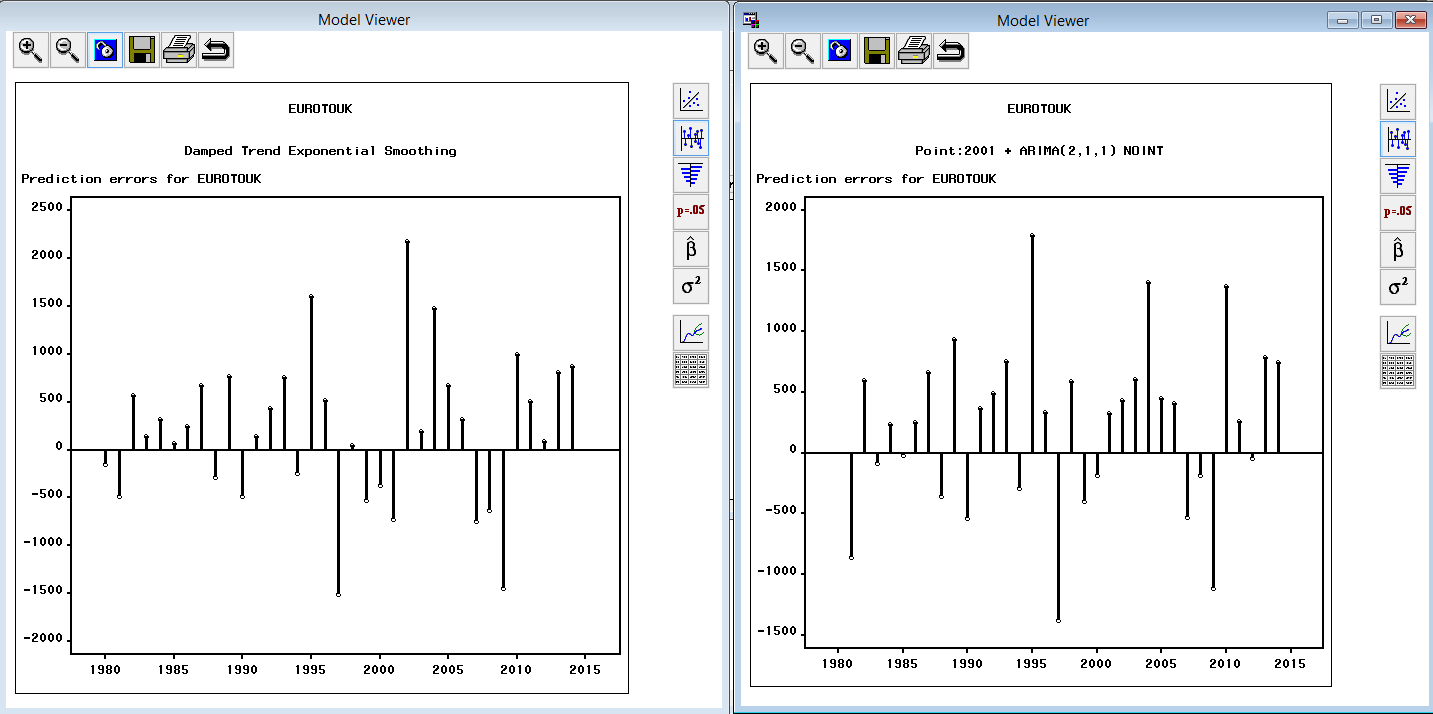
**Residents from EUROPE Visiting UK**



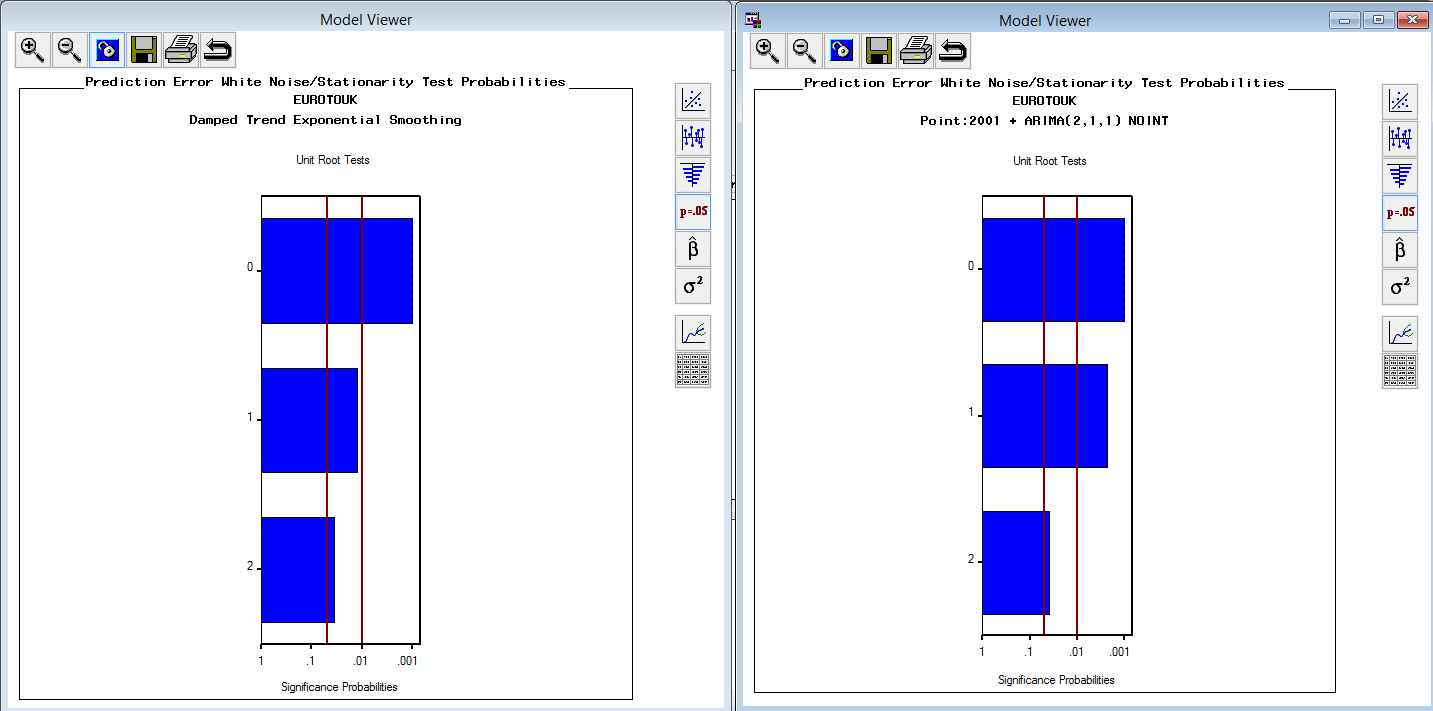
Many model combinations were tried, among which ARIMA (2,1,1) with point intervention is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

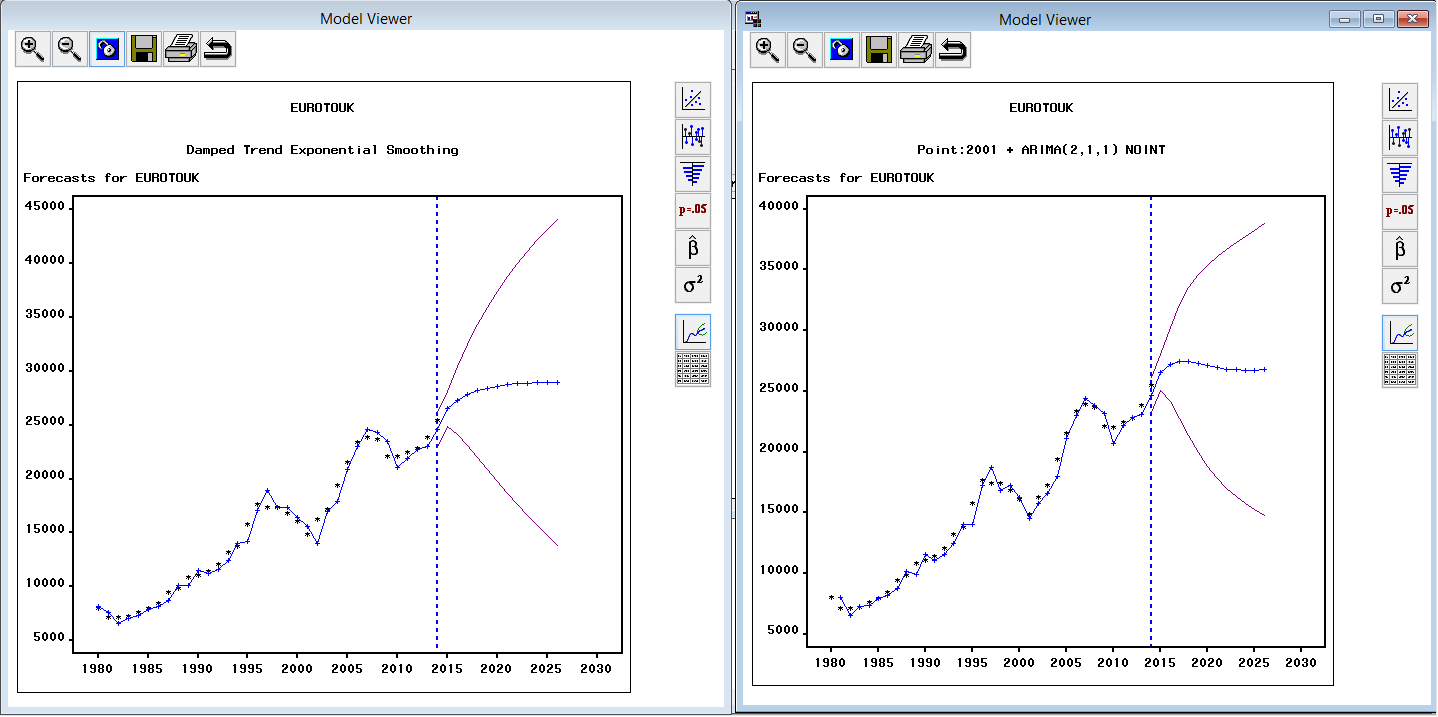




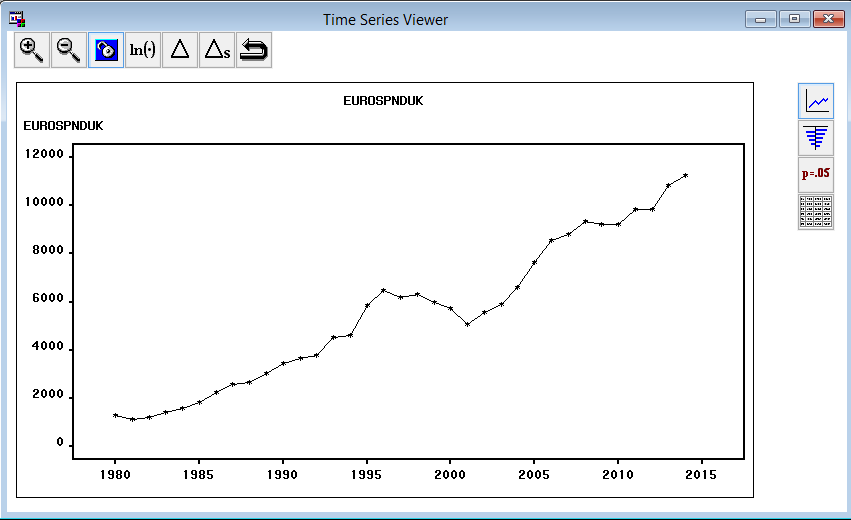




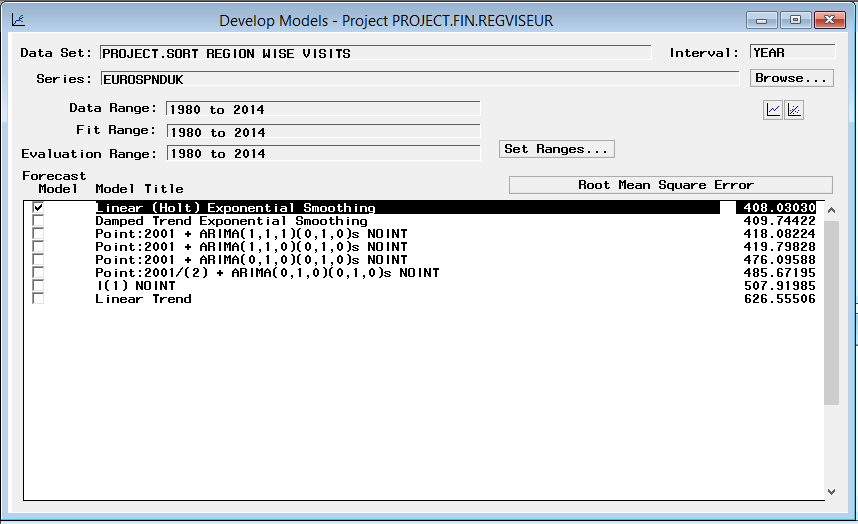


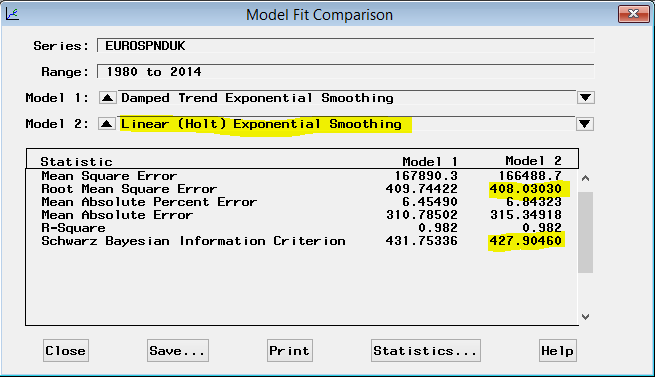


**Residents from Europe spending in UK**

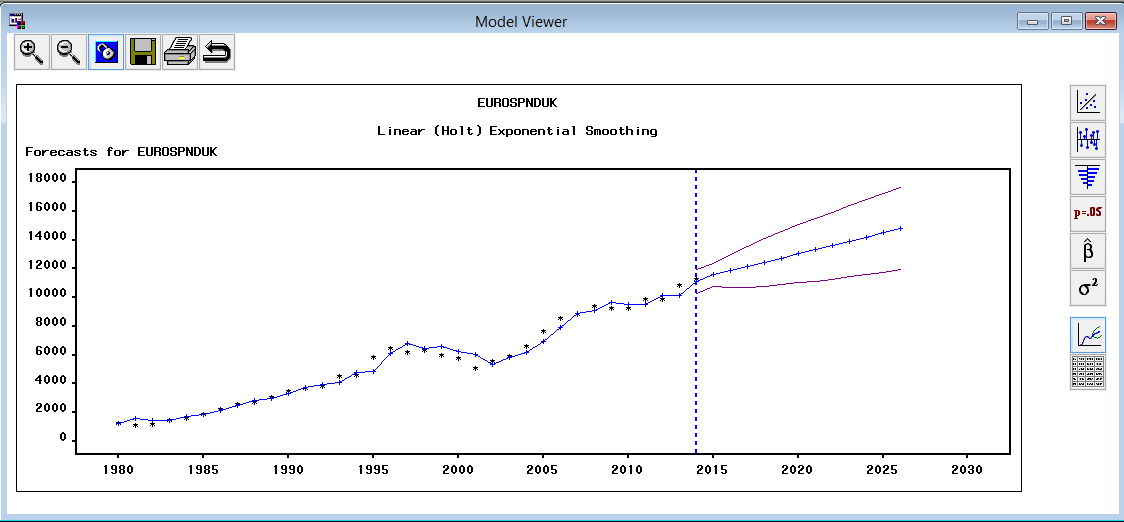


Many model combinations were tried, among which Linear (holt) Exponential Smoothing, built with just simple differencing is selected as the best fitting model after examining ACF, PACF and IACF plots etc.



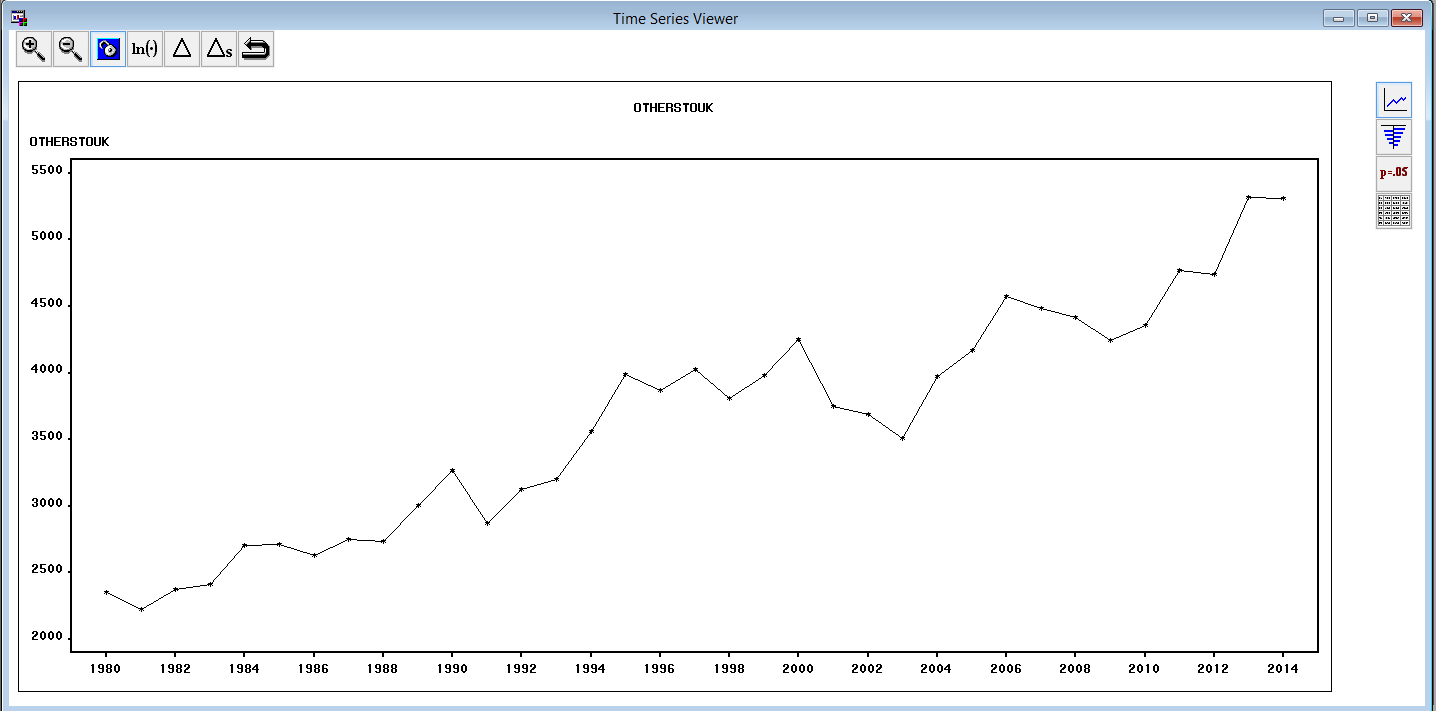


Linear (holt) exponential smoothing gives us the model with very low confidence interval which is good.

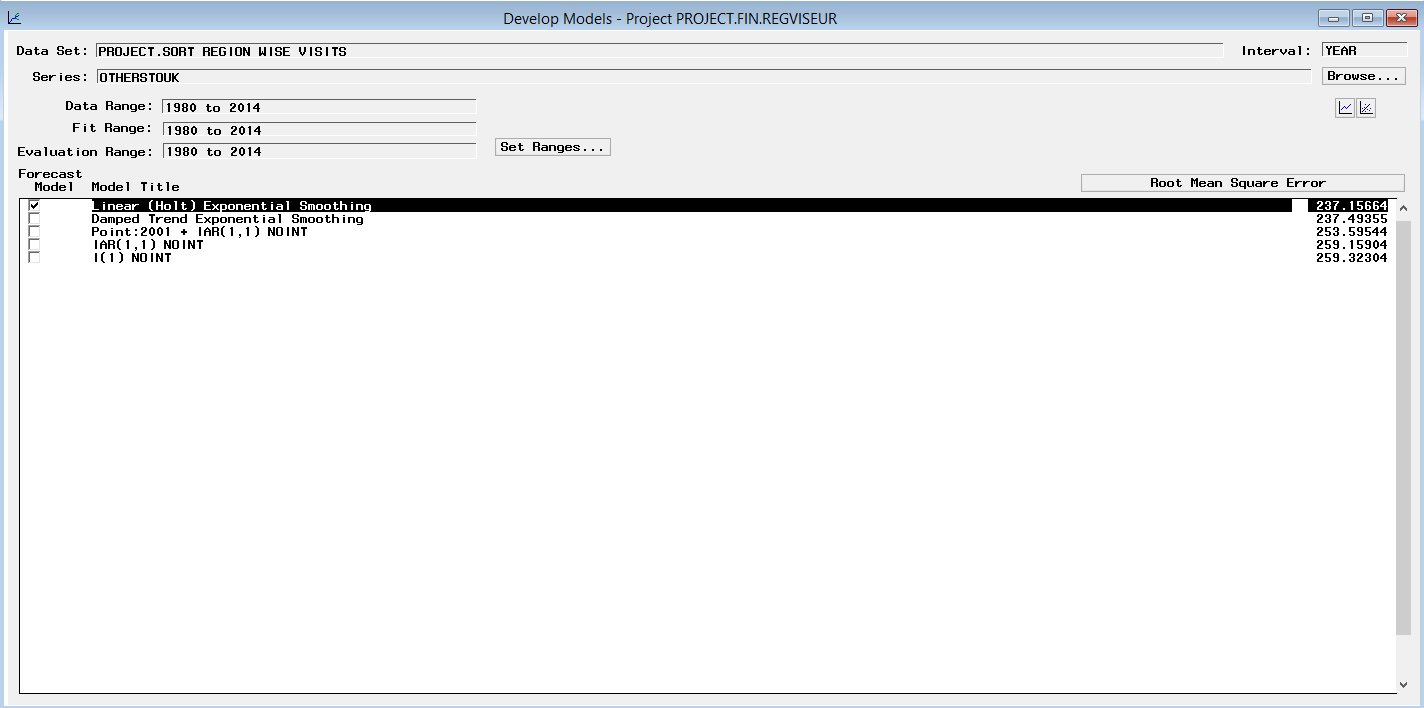


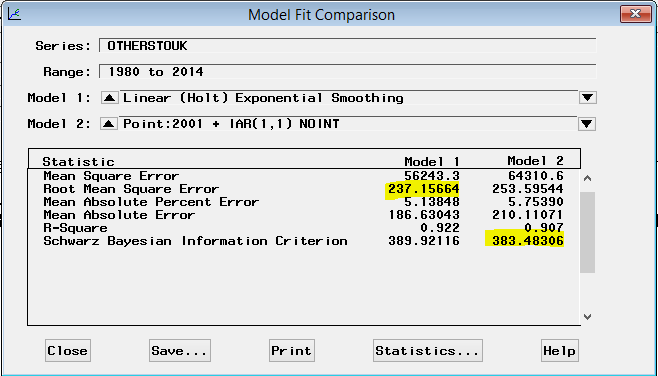
Hence, we go with this model as the best fit.

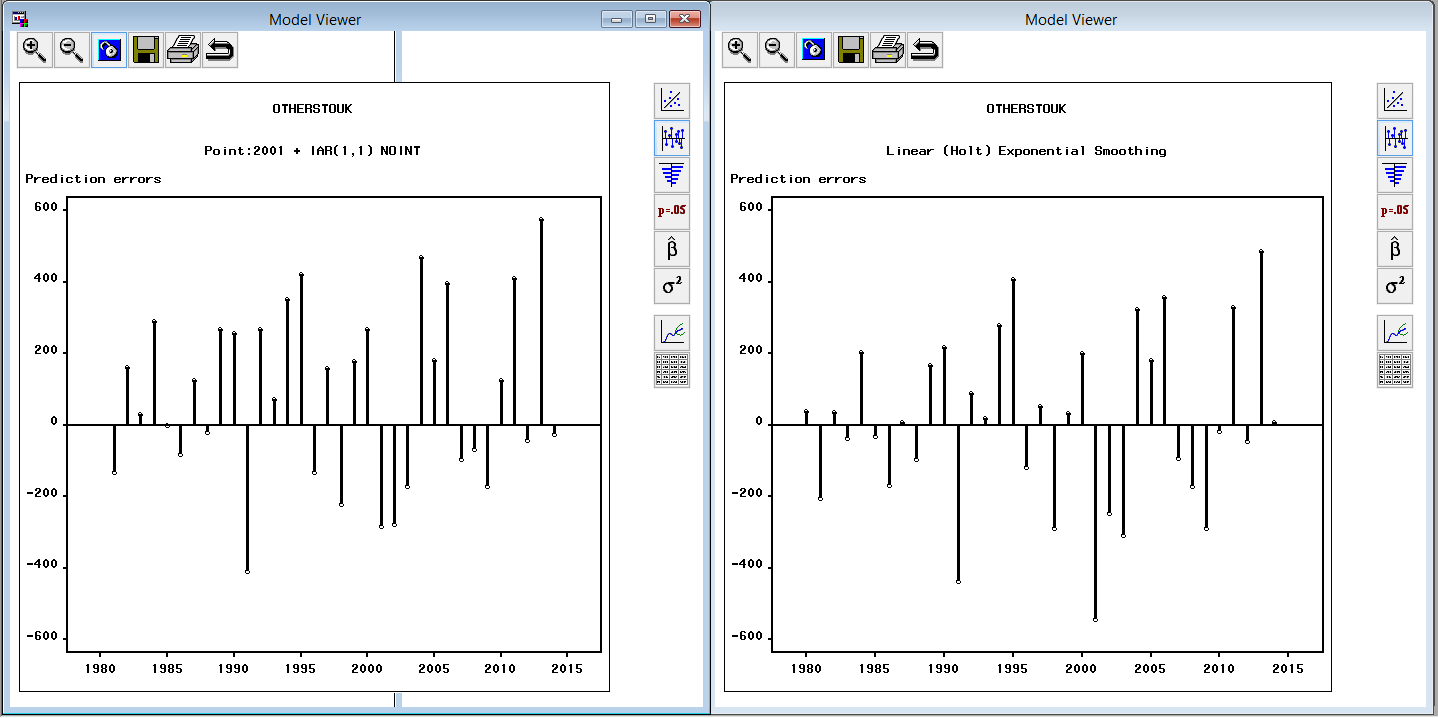
**Residents from Other regions visiting UK**

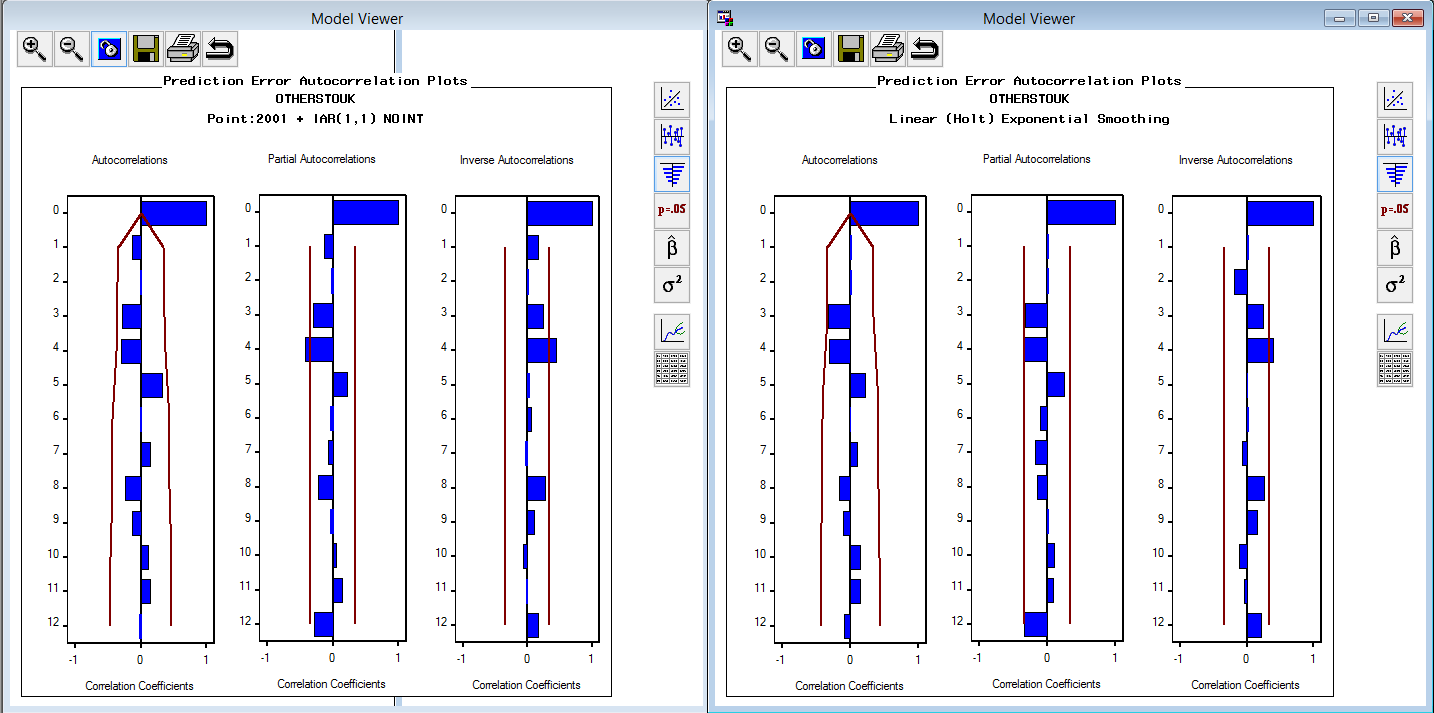


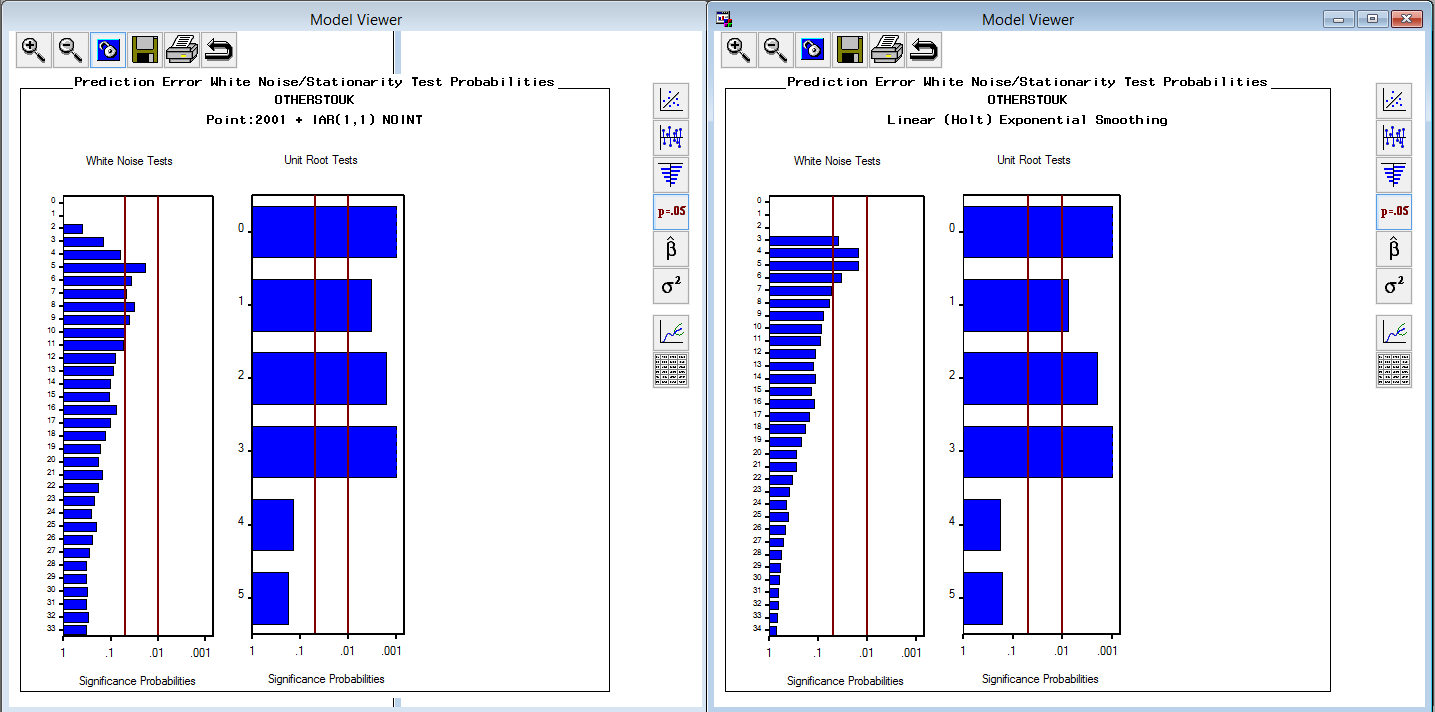
Many model combinations were tried, among which Linear (Holt) Exponential Smoothing, built with just simple differencing is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

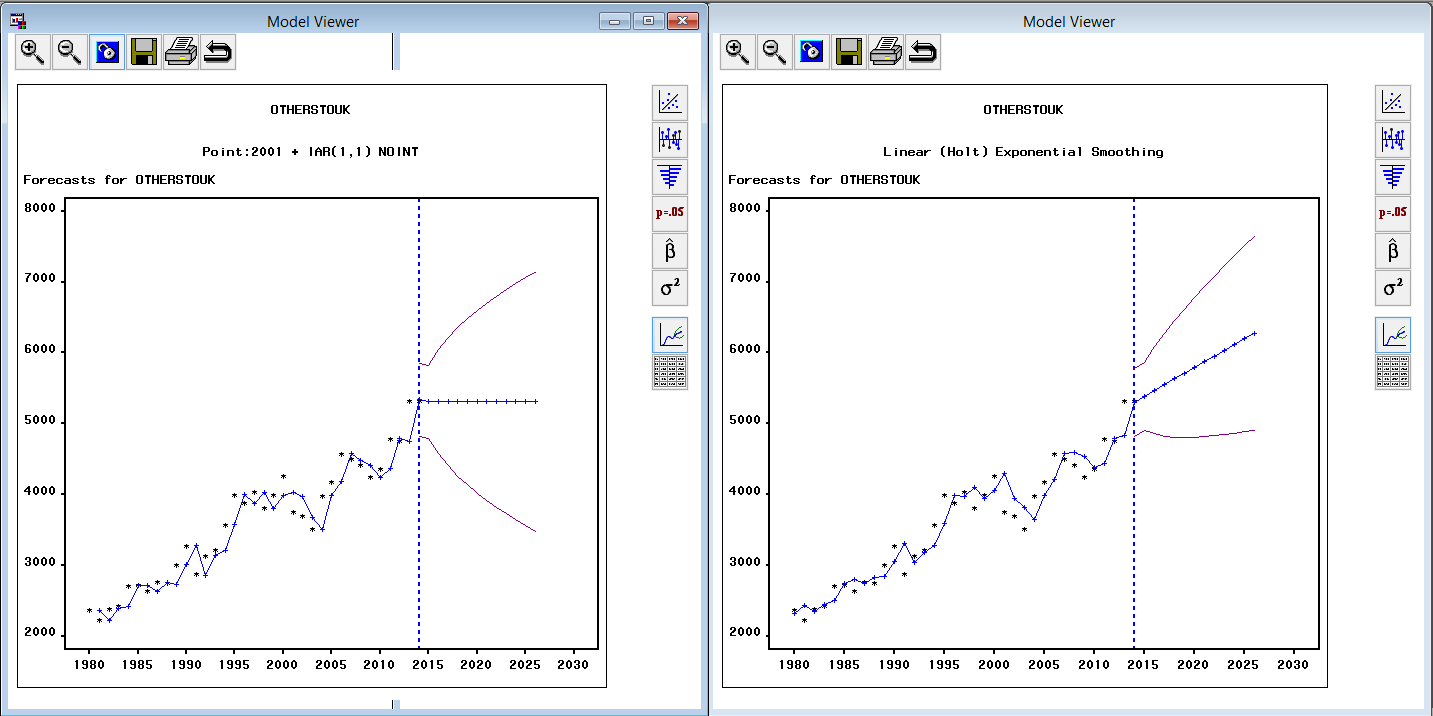




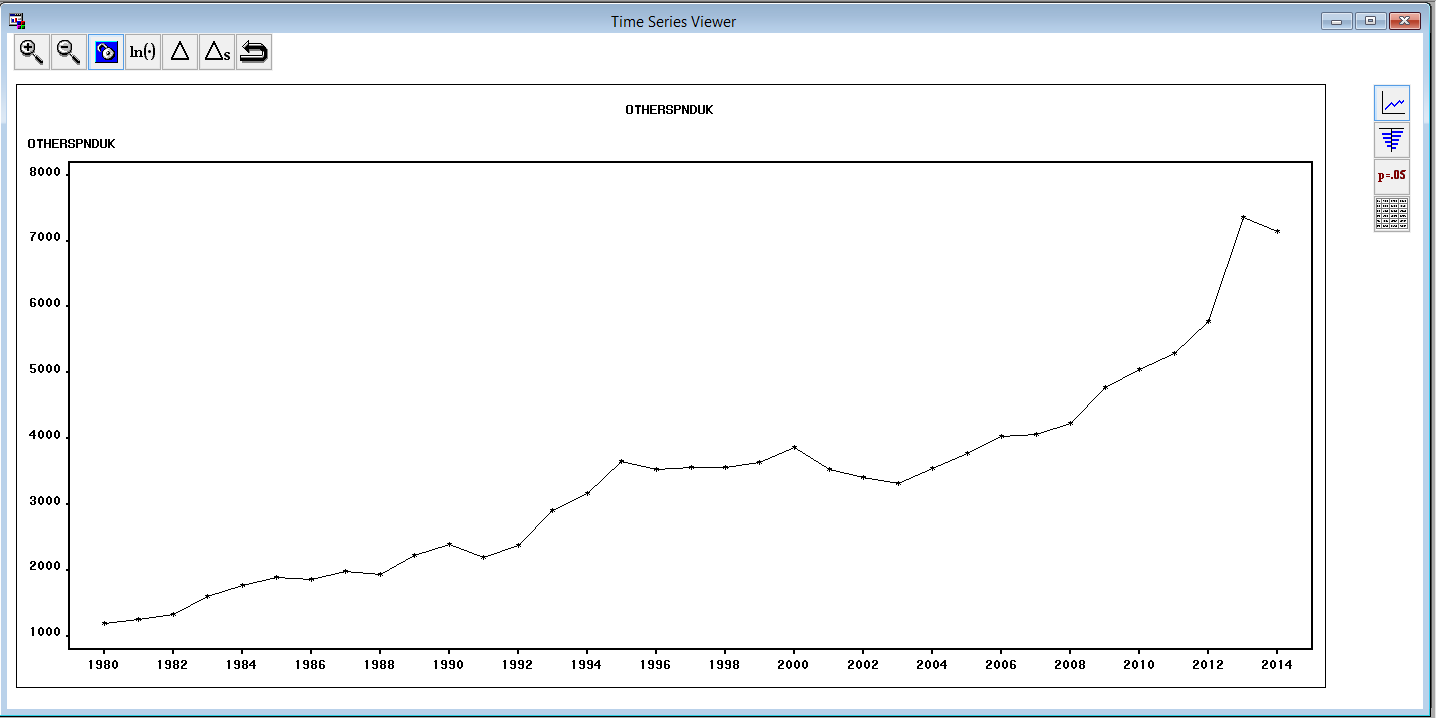




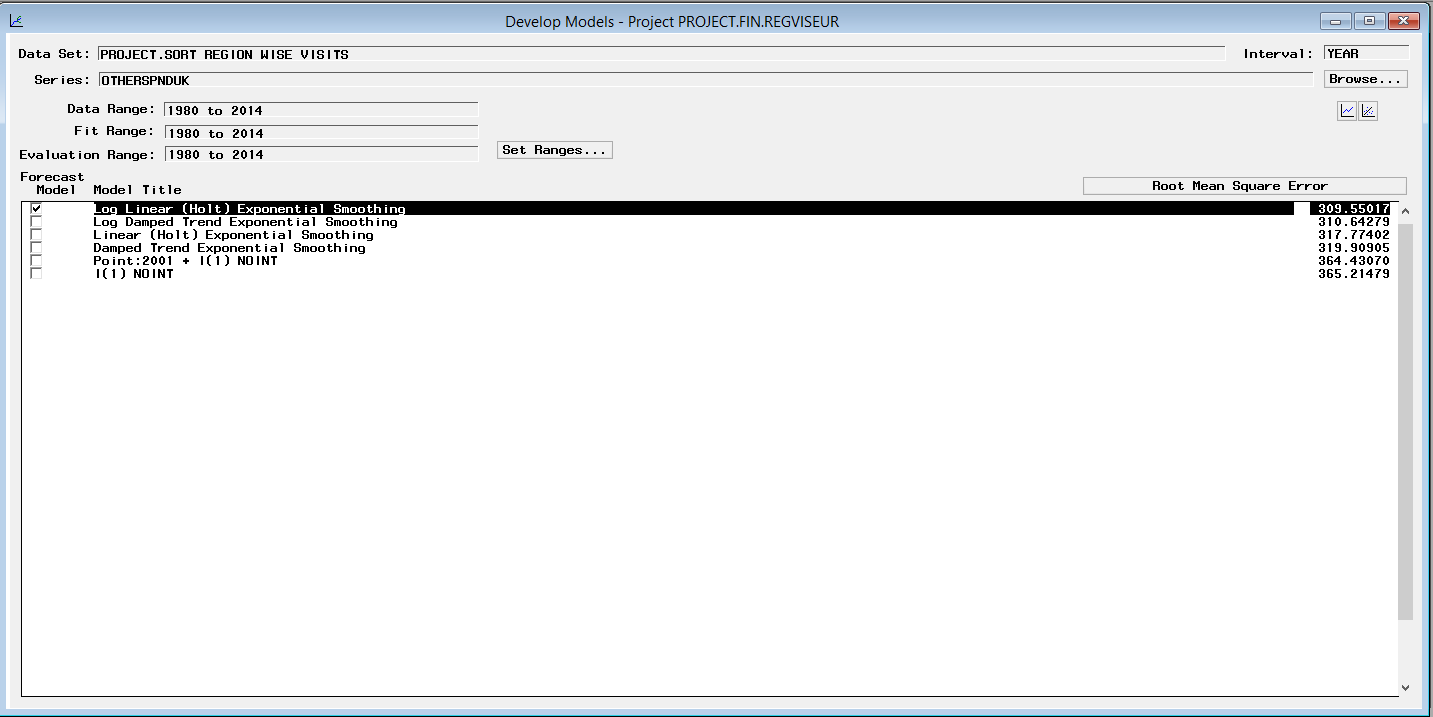


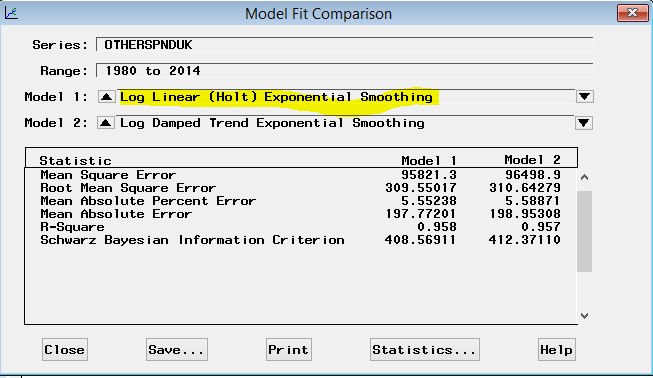


**Residents from Other regions spending in UK**

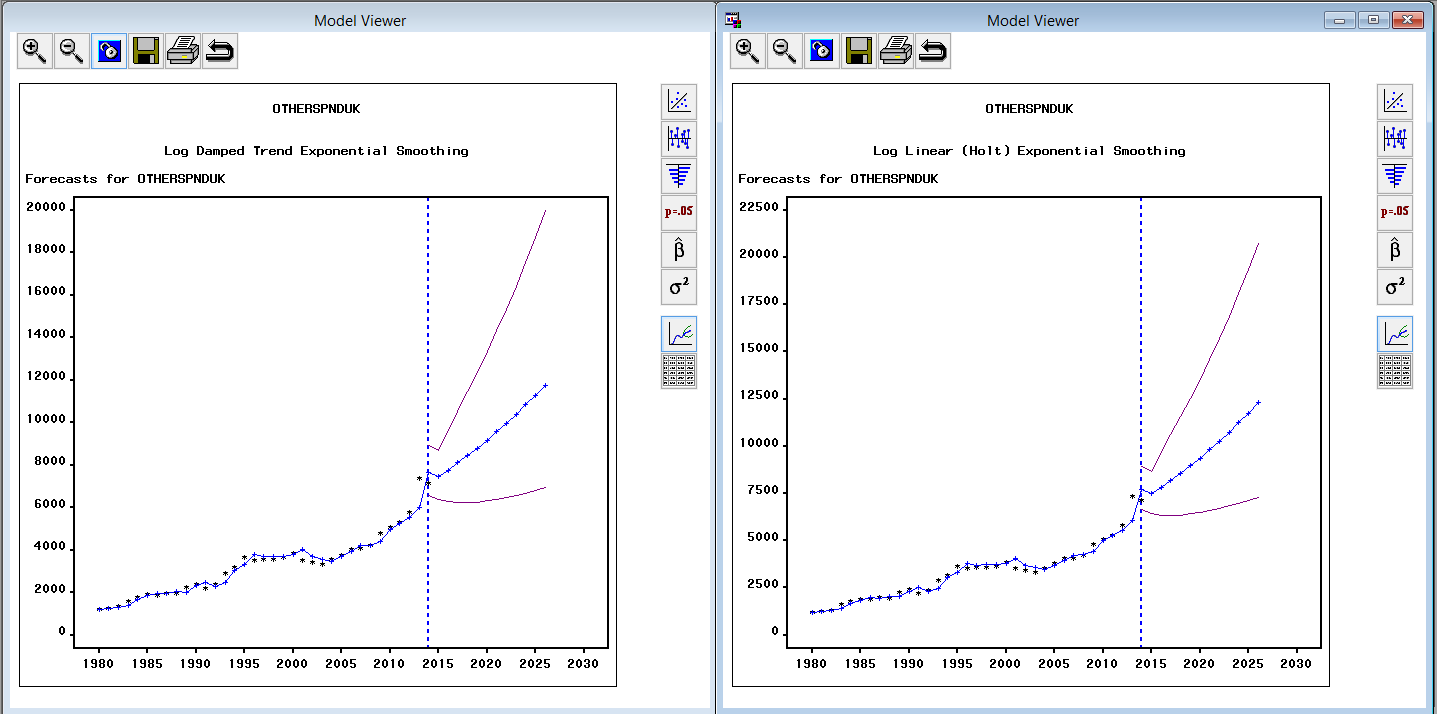


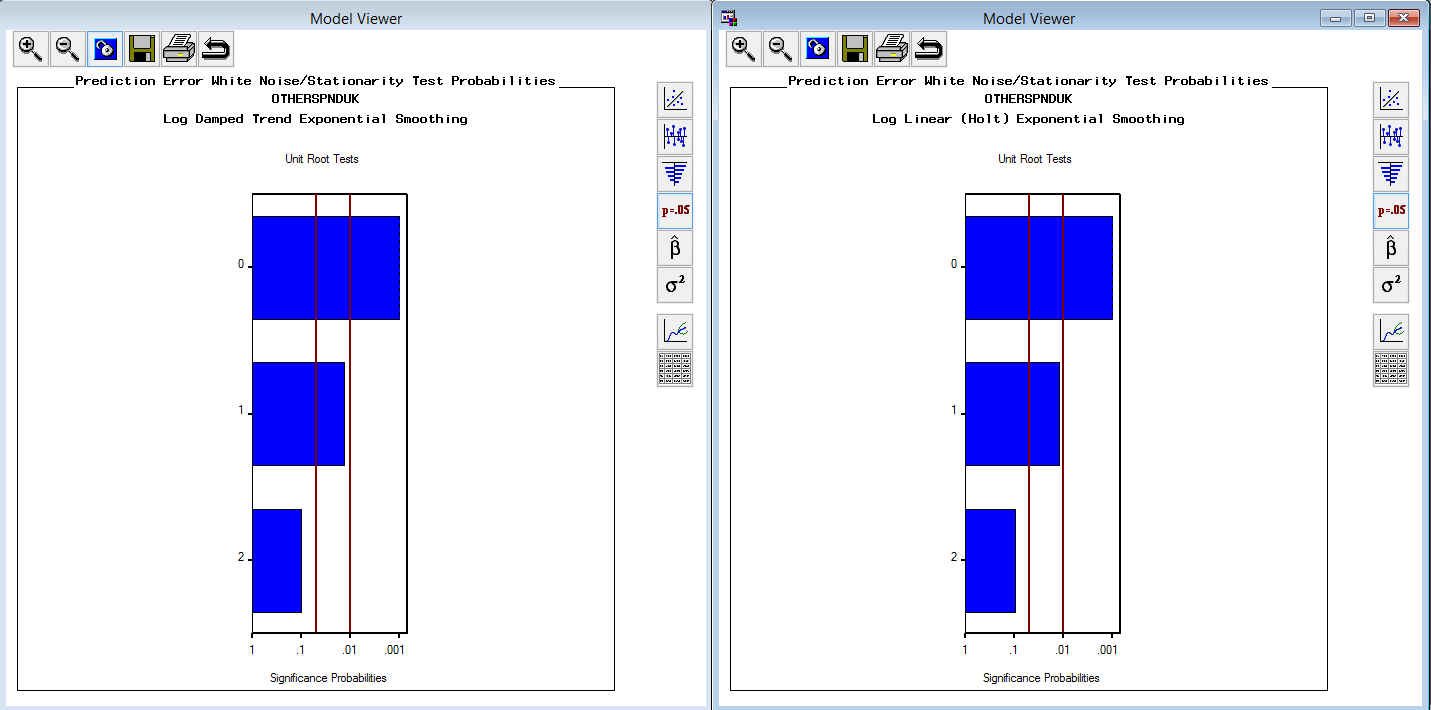
Many model combinations were tried, among which Log Linear (holt) Exponential Smoothing, built with just simple differencing is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

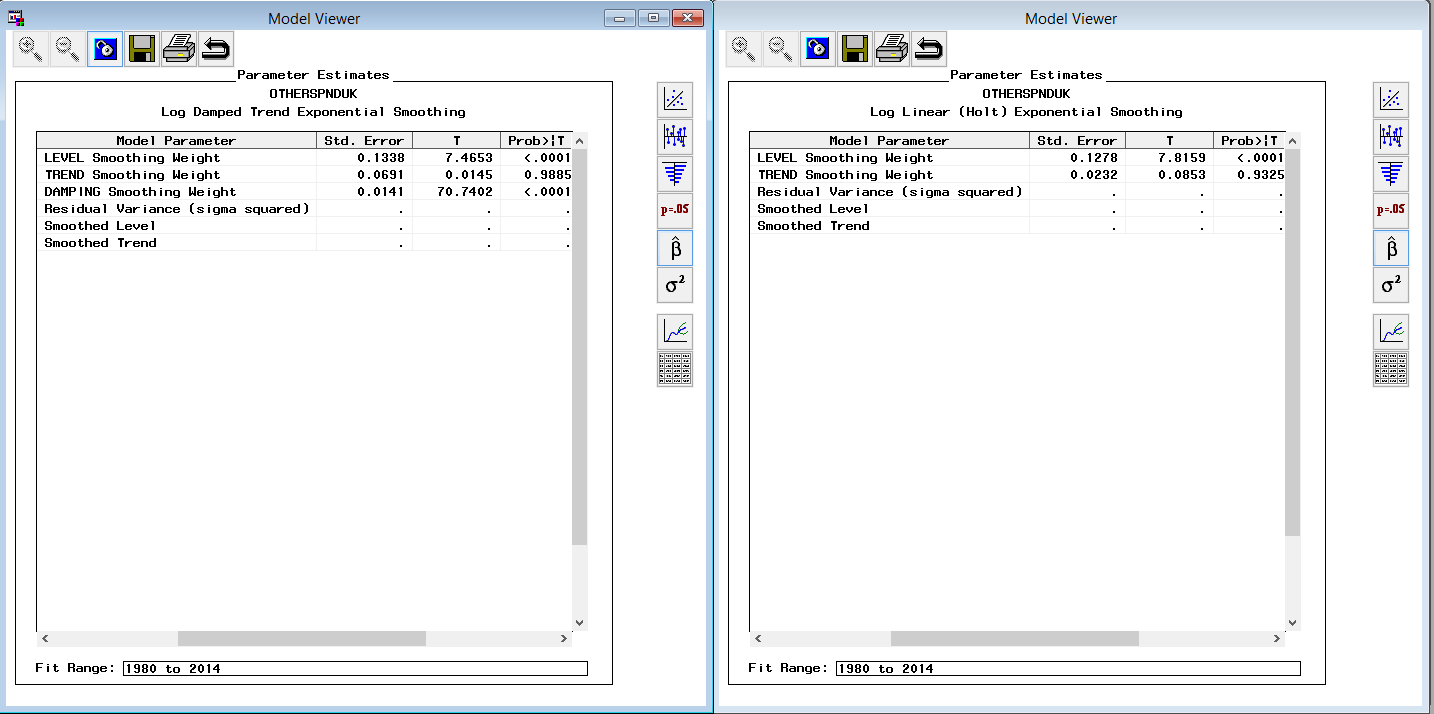




No difference as such as seen below:







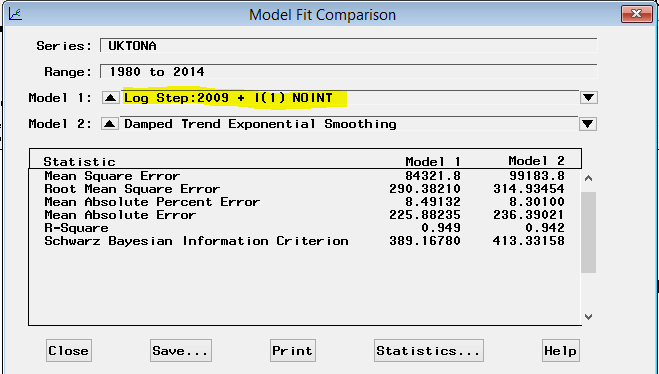
**UK Residents visiting North America**

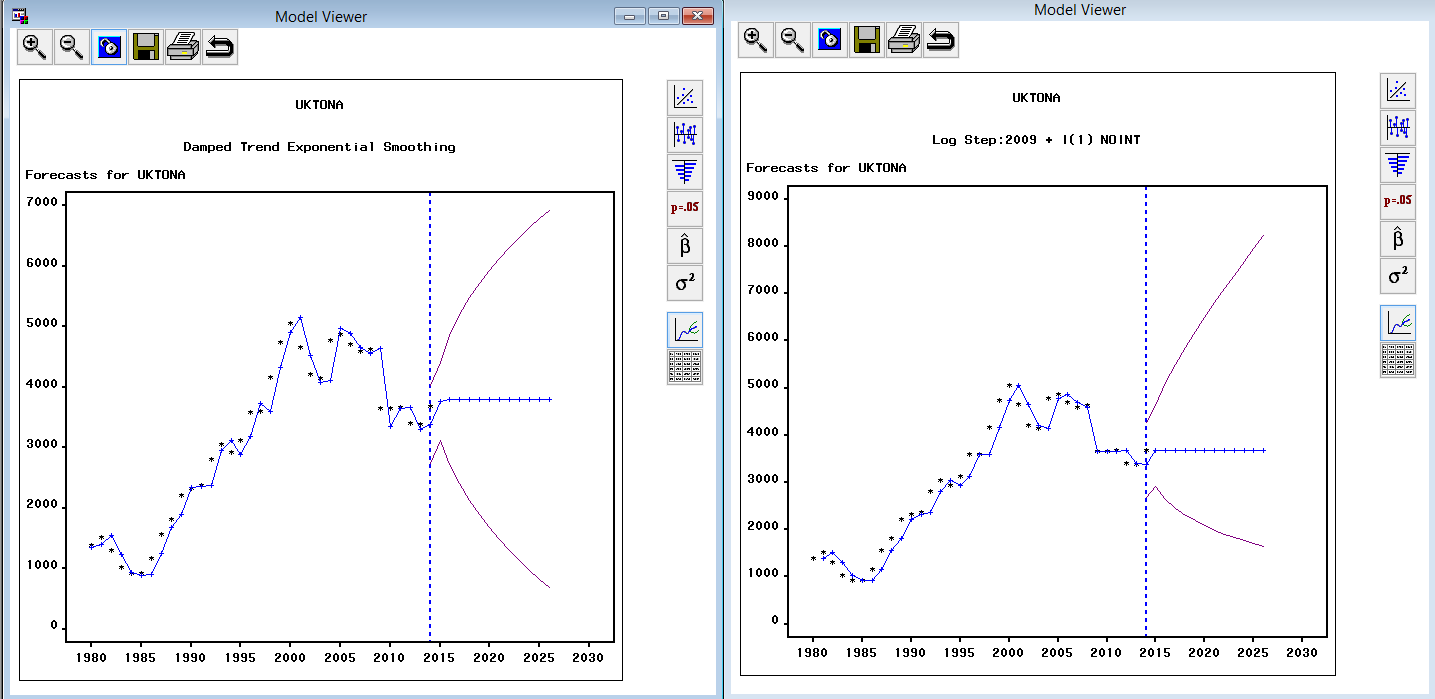


Many model combinations were tried, among which 1(1), built with simple differencing and step intervention with log transformation is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

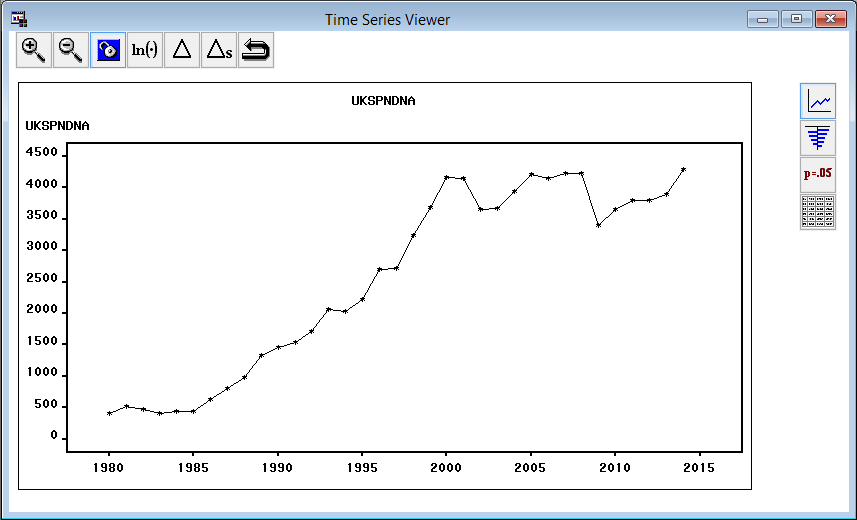


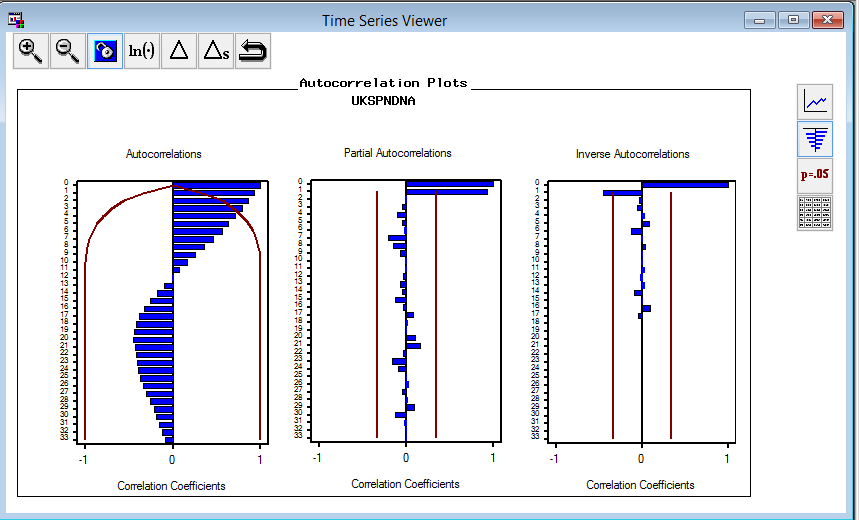
Comparison between two models

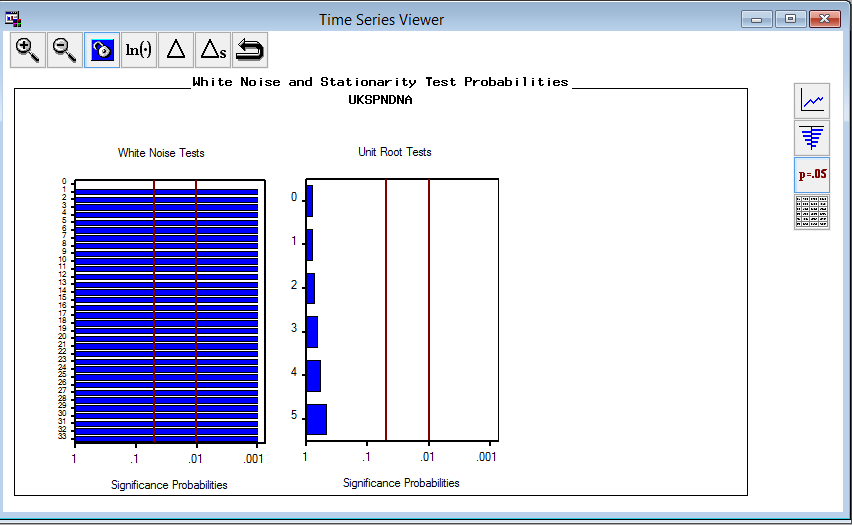




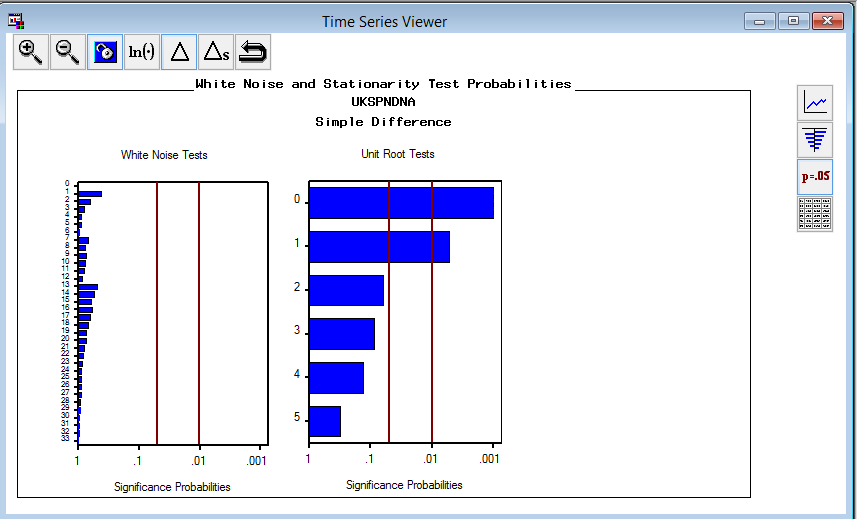
**UK Residents Spending in North America**

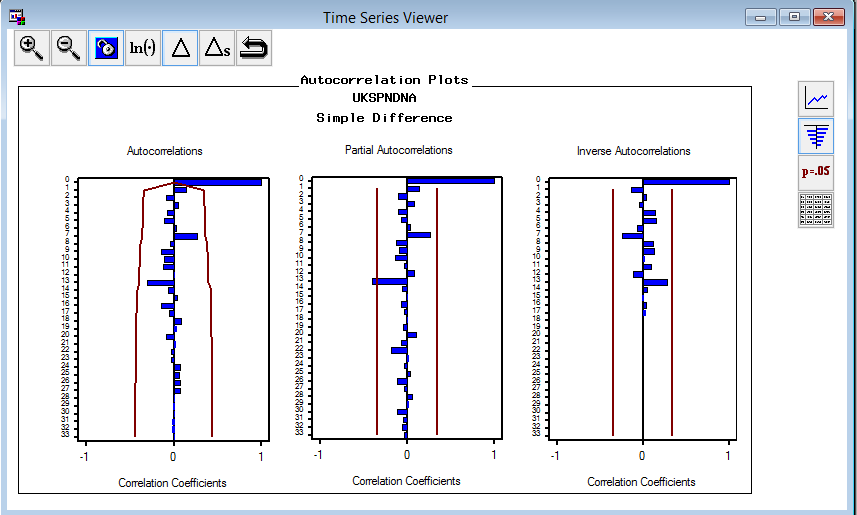




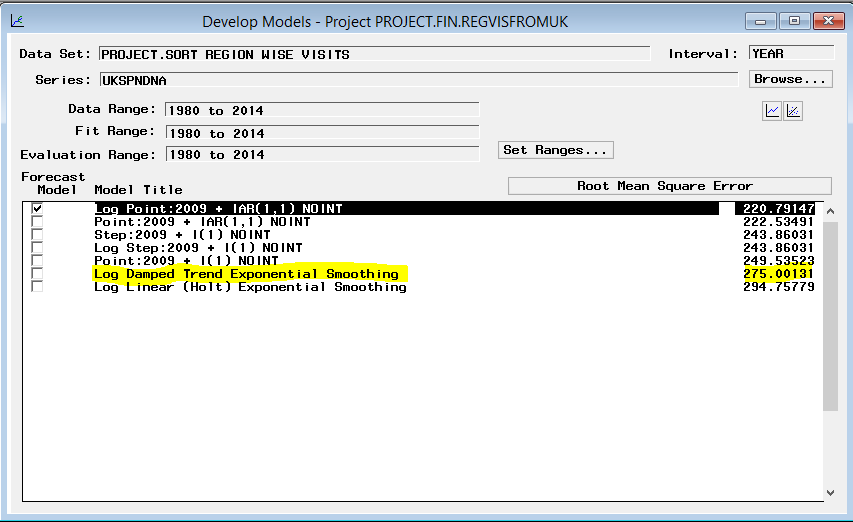


After first differencing,



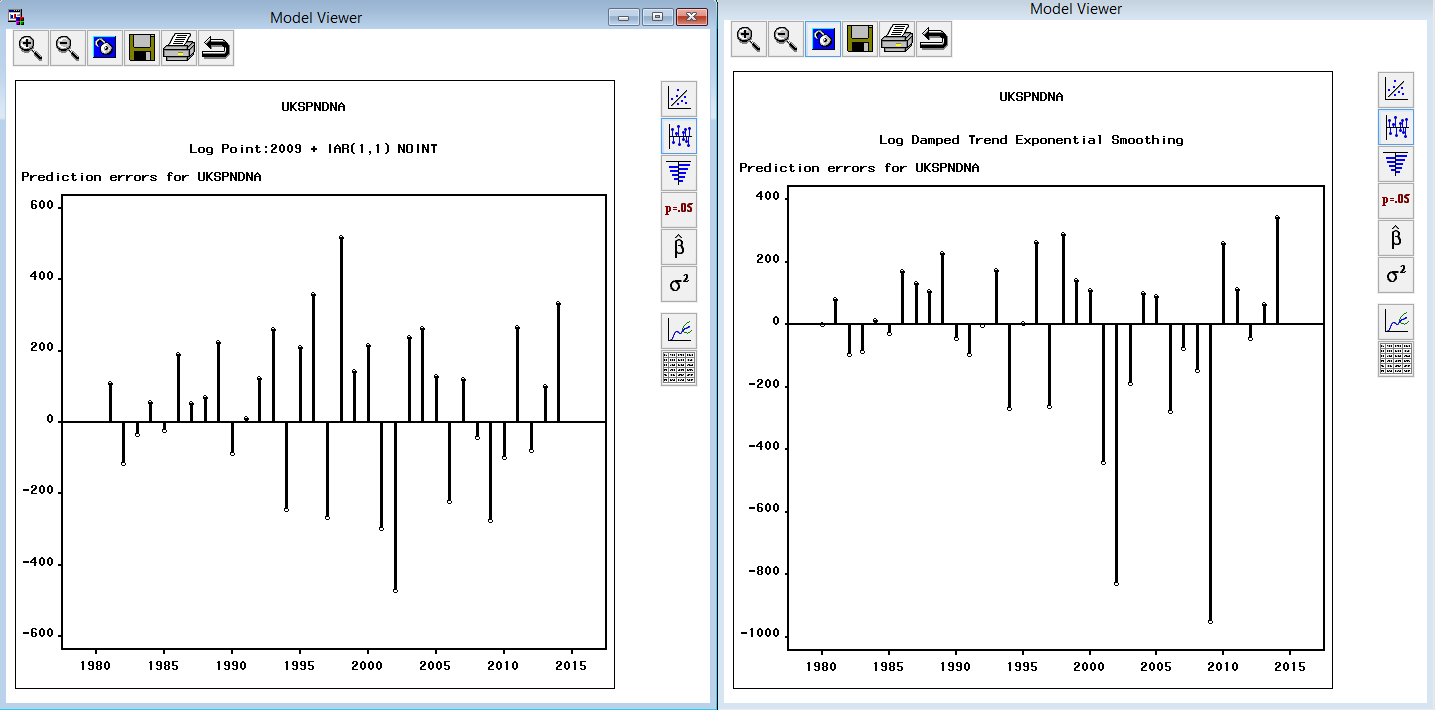


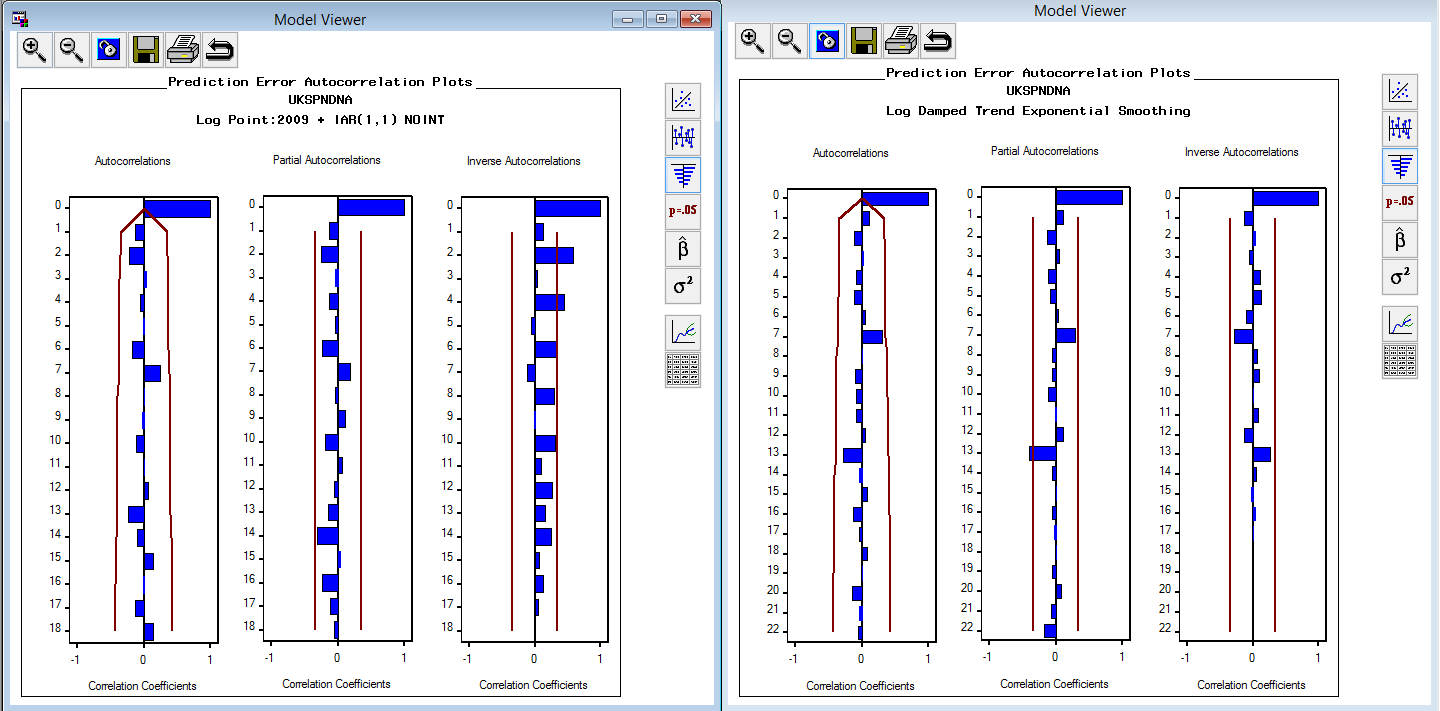
Many model combinations were tried, among which IAR (1,1) with point intervention and log transformation is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

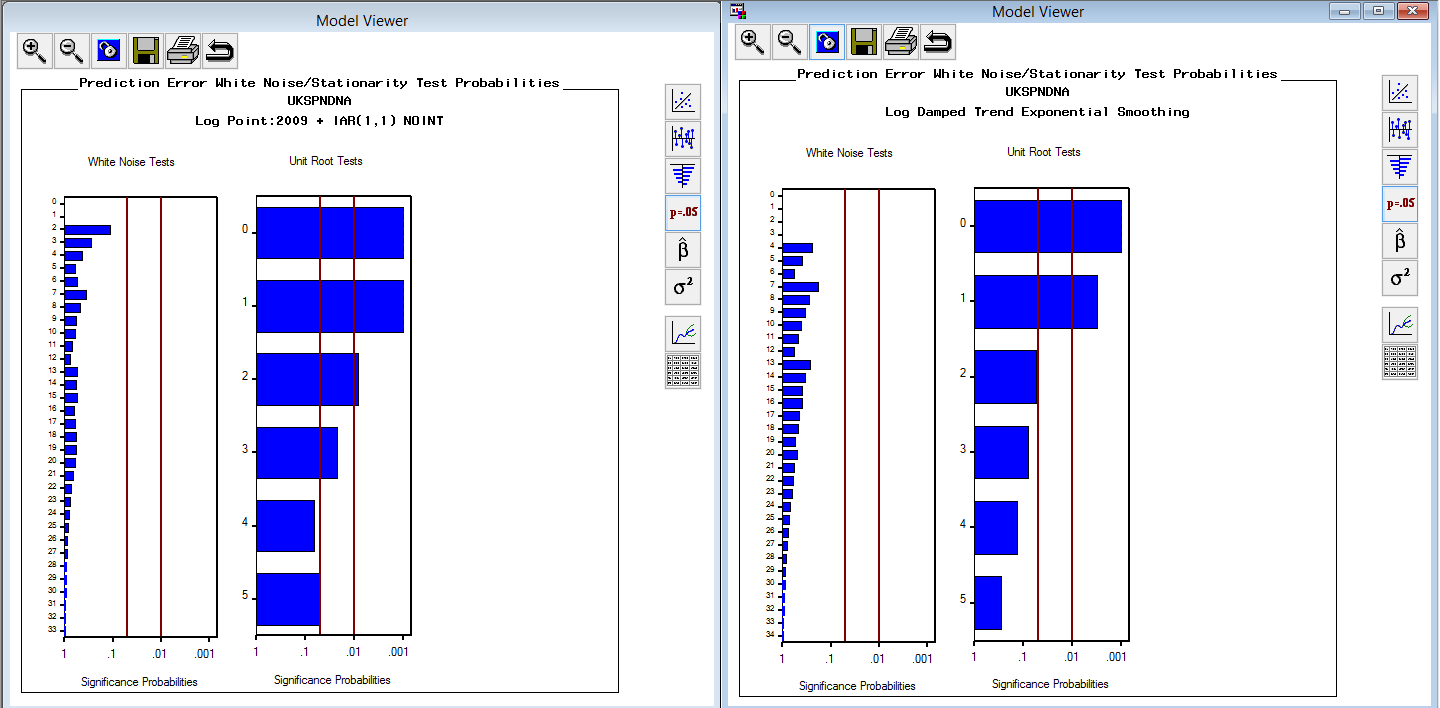


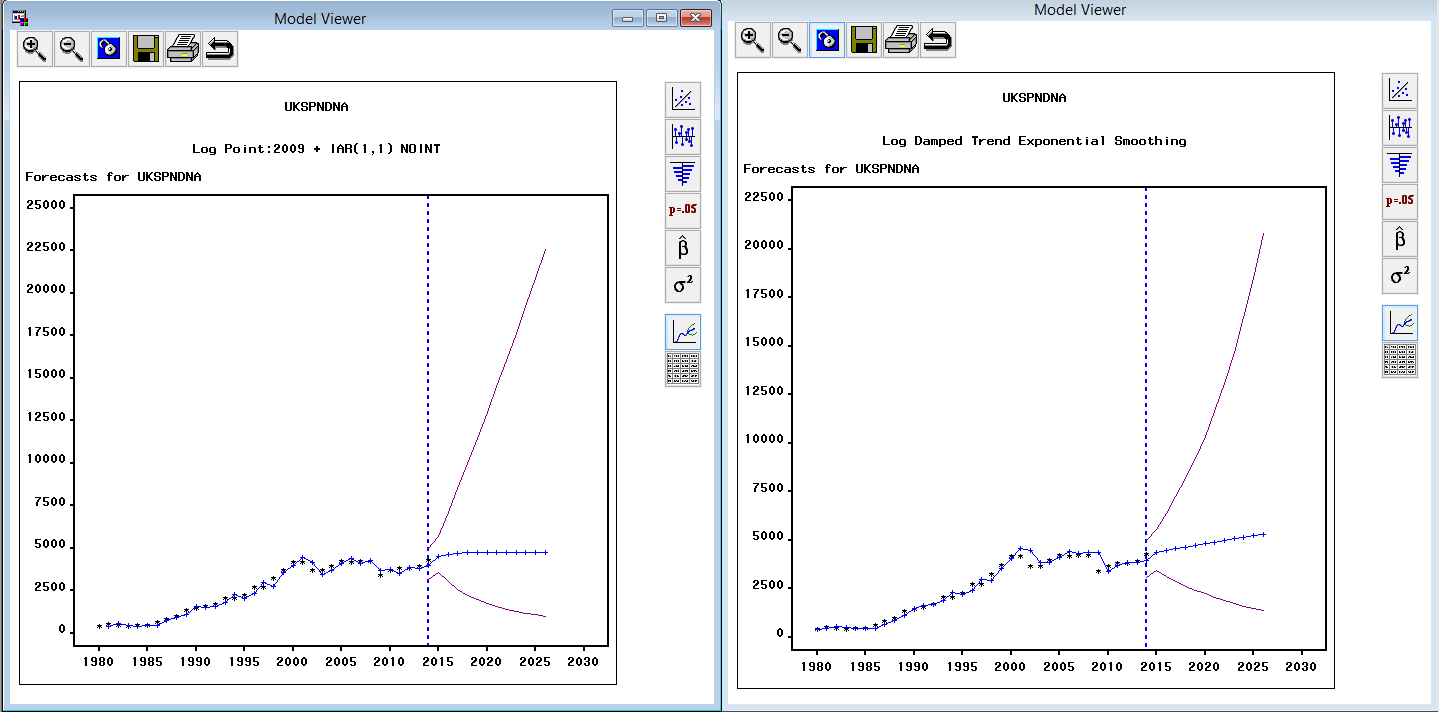


Comparison between two models

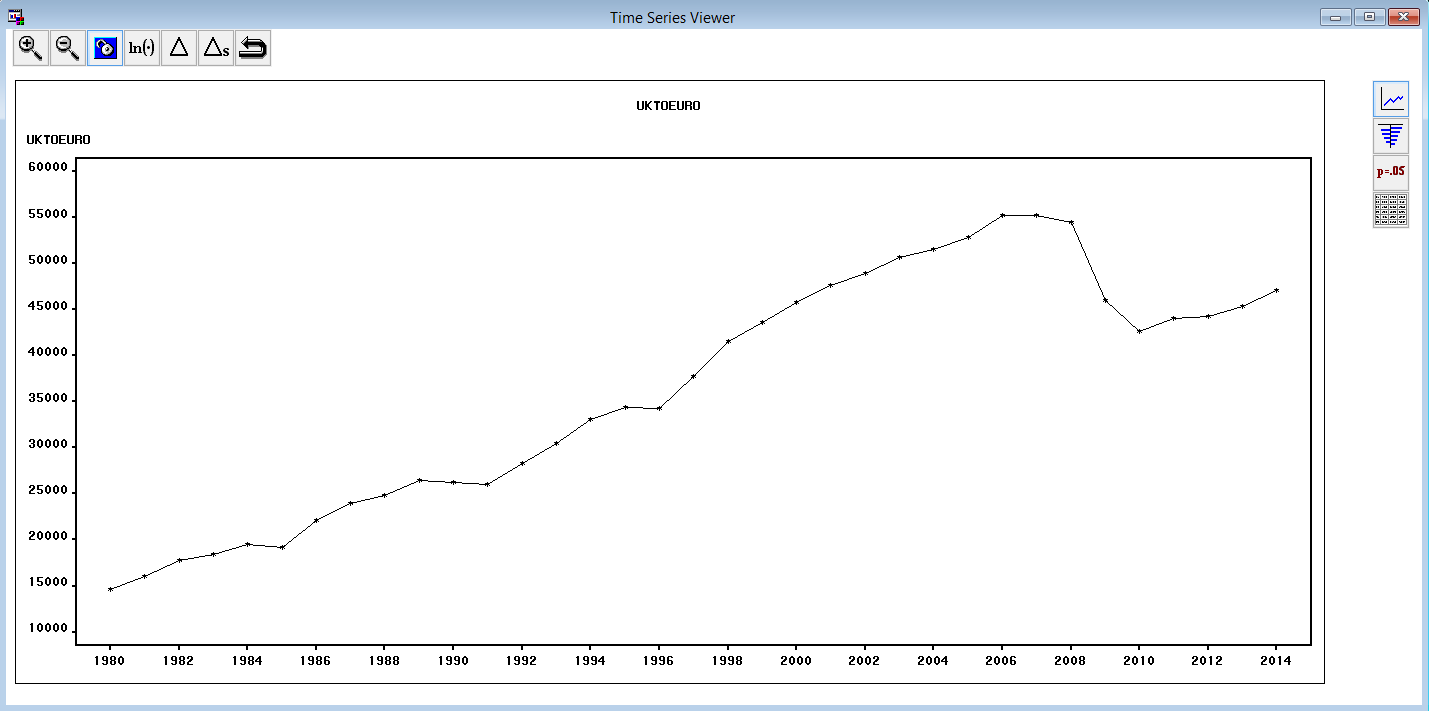




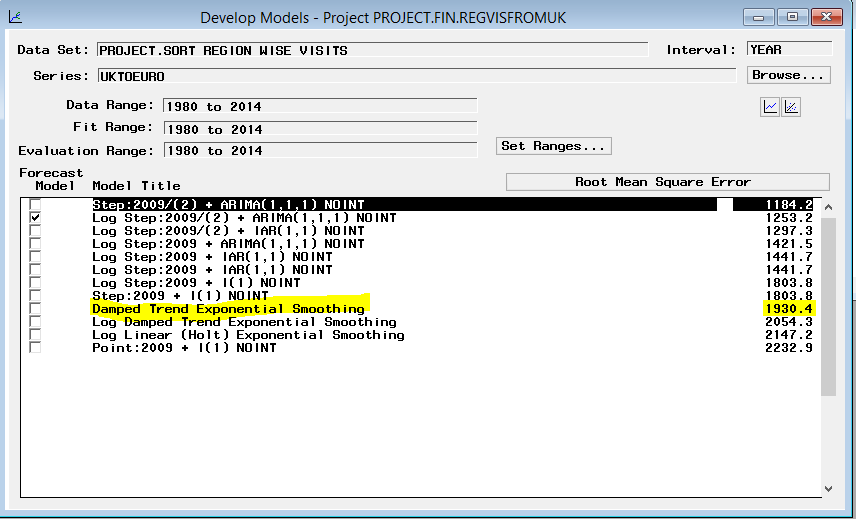


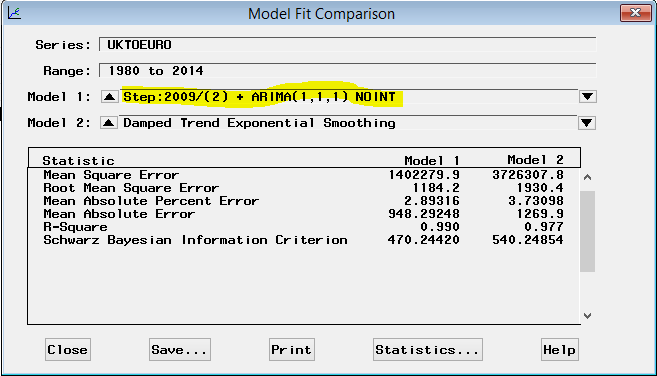


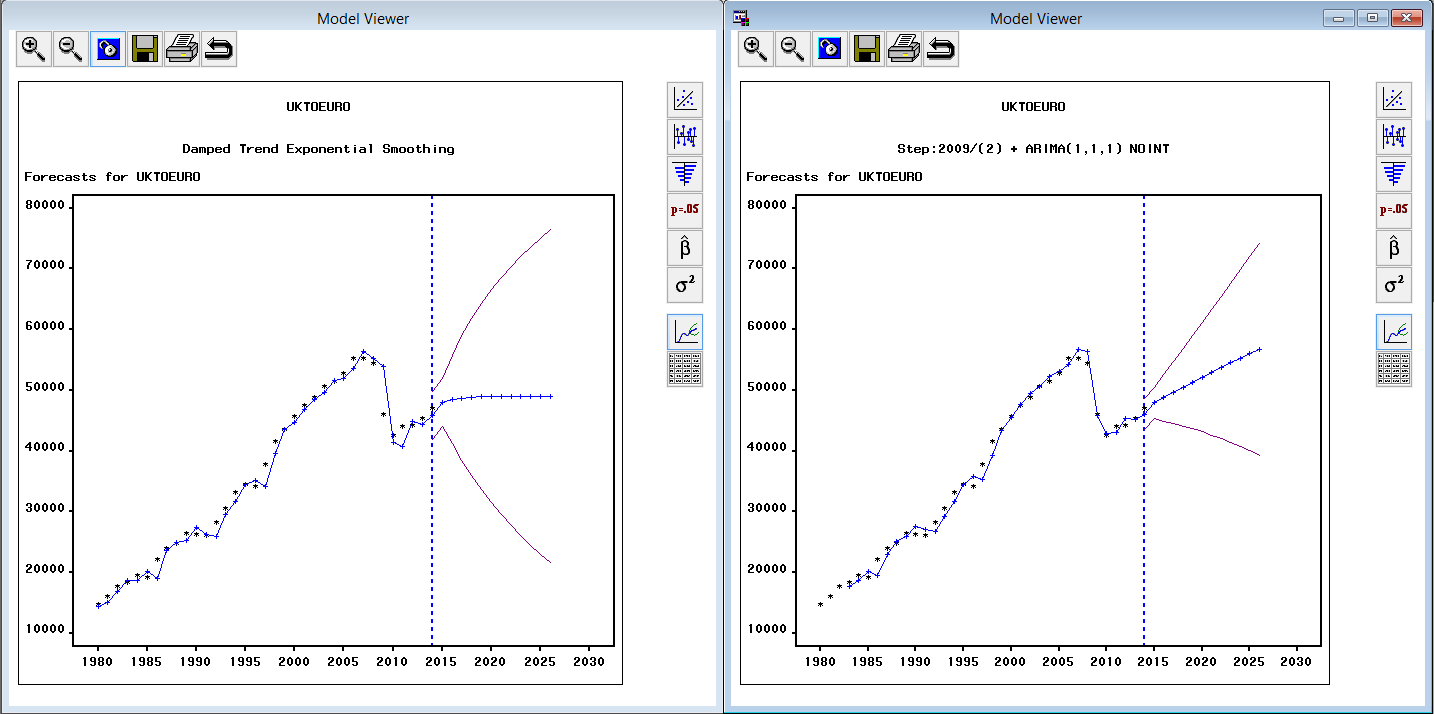
**UK Residents visiting EUROPE**



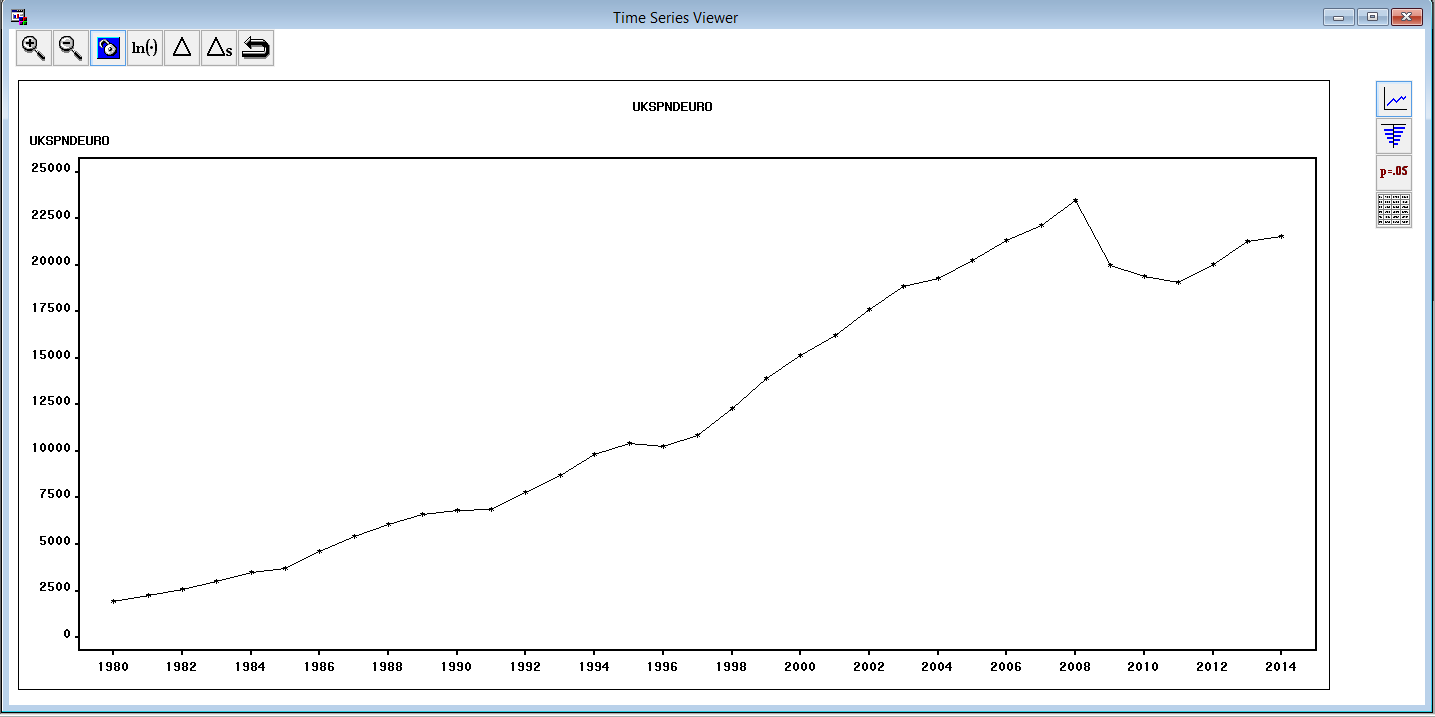
Many model combinations were tried, among which ARIMA(1,1,1) with STEP intervention along with WAVE is selected as the best fitting model after examining ACF, PACF and IACF plots etc.



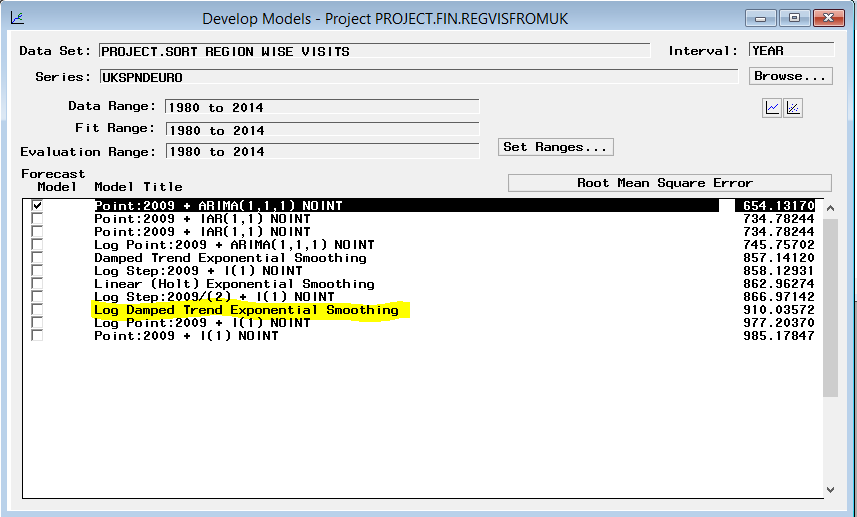


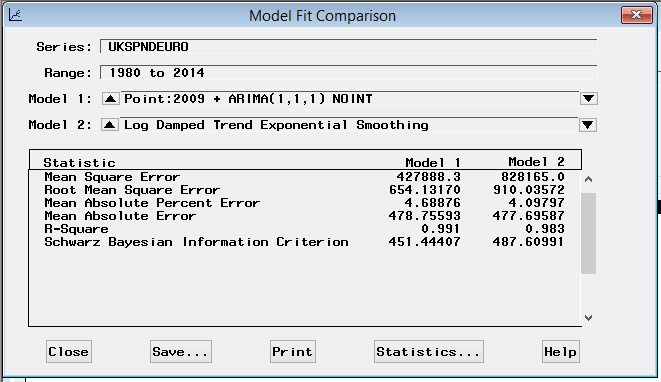


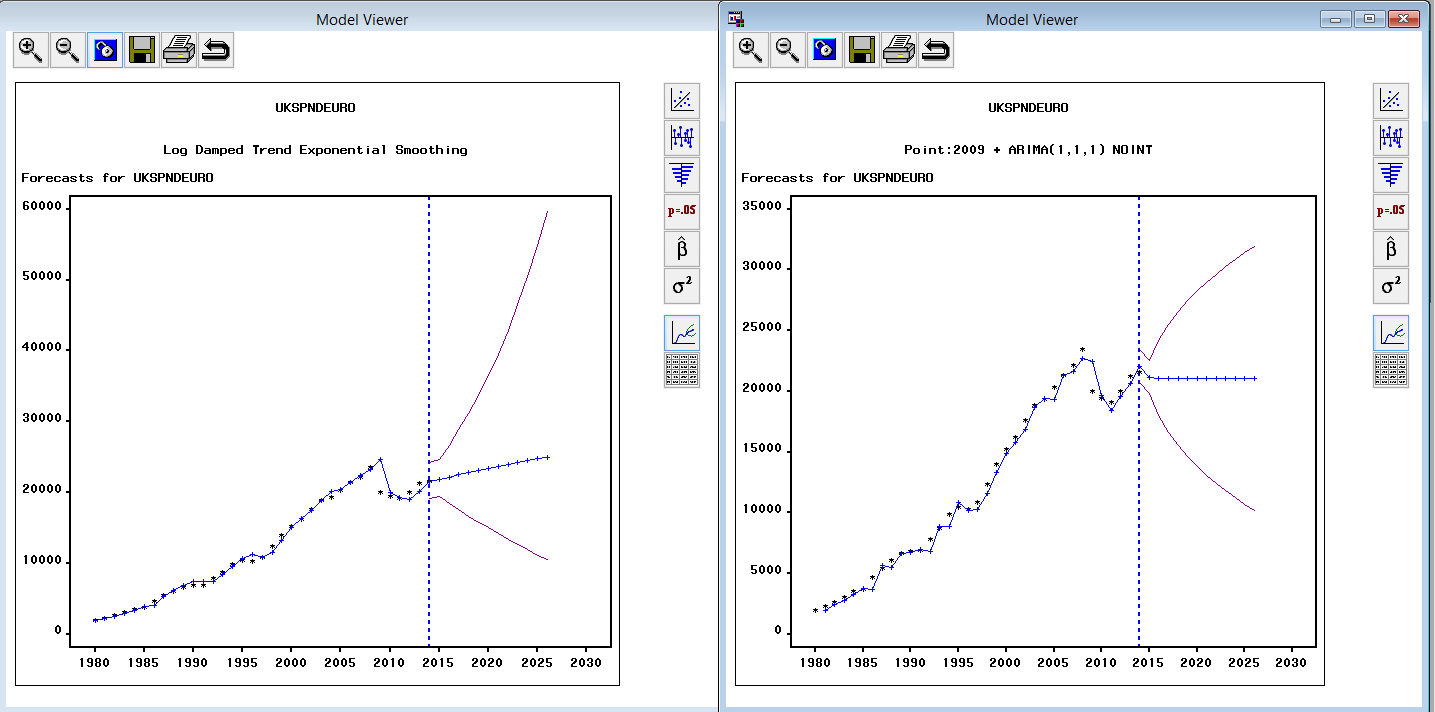
**UK Residents Spending’s in EUROPE**



Many model combinations were tried, among which ARIMA (1,1,1) with POINT intervention is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

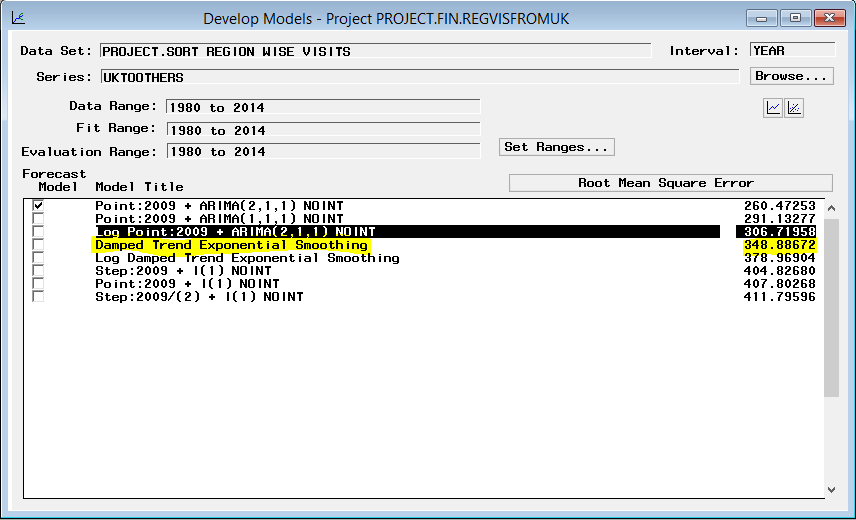


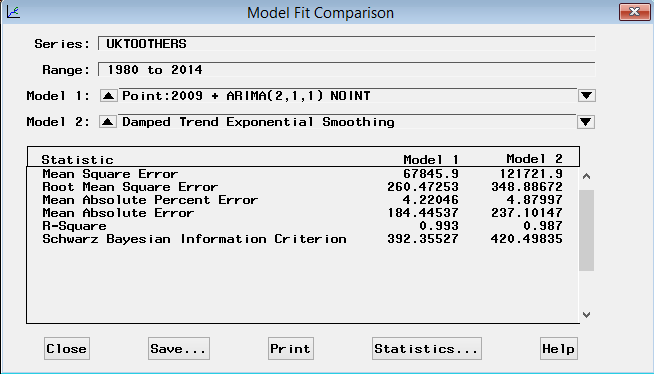


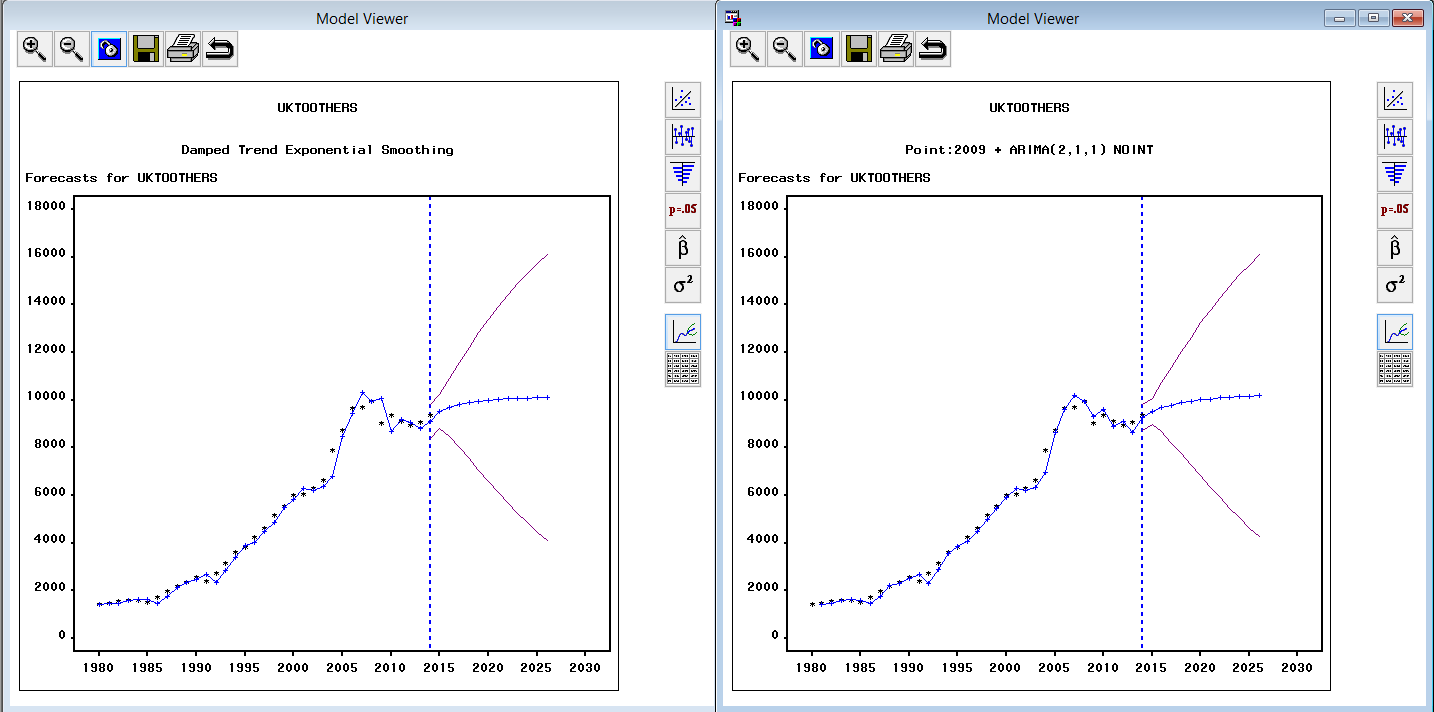


**UK Residents visiting Other regions**

Many model combinations were tried, among which ARIMA (2,1,1) with POINT intervention is selected as the best fitting model after examining ACF, PACF and IACF plots etc.

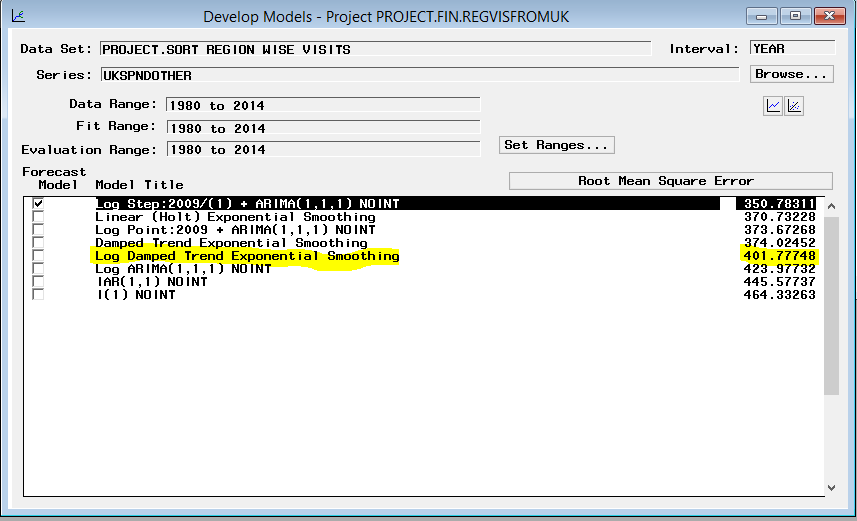


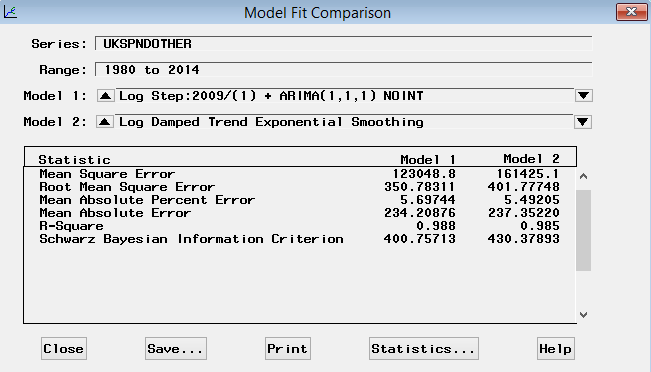


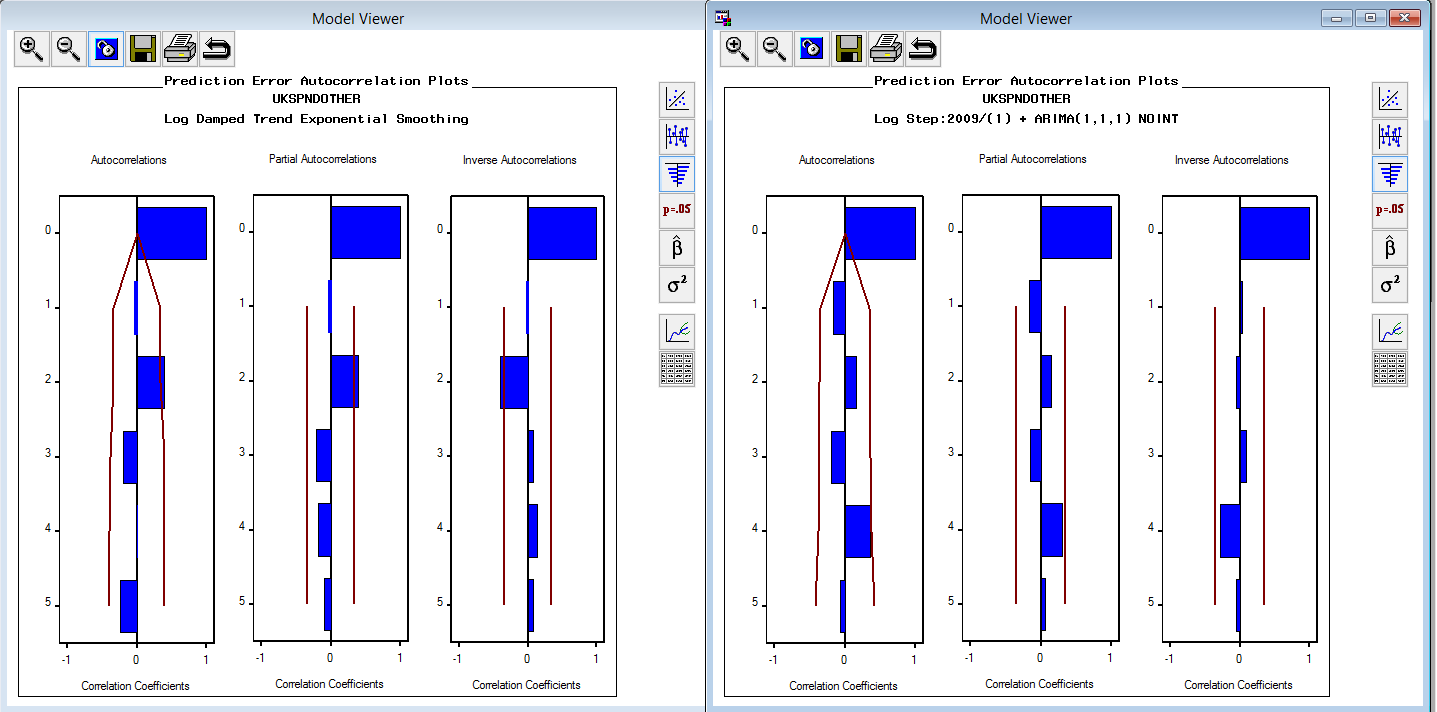


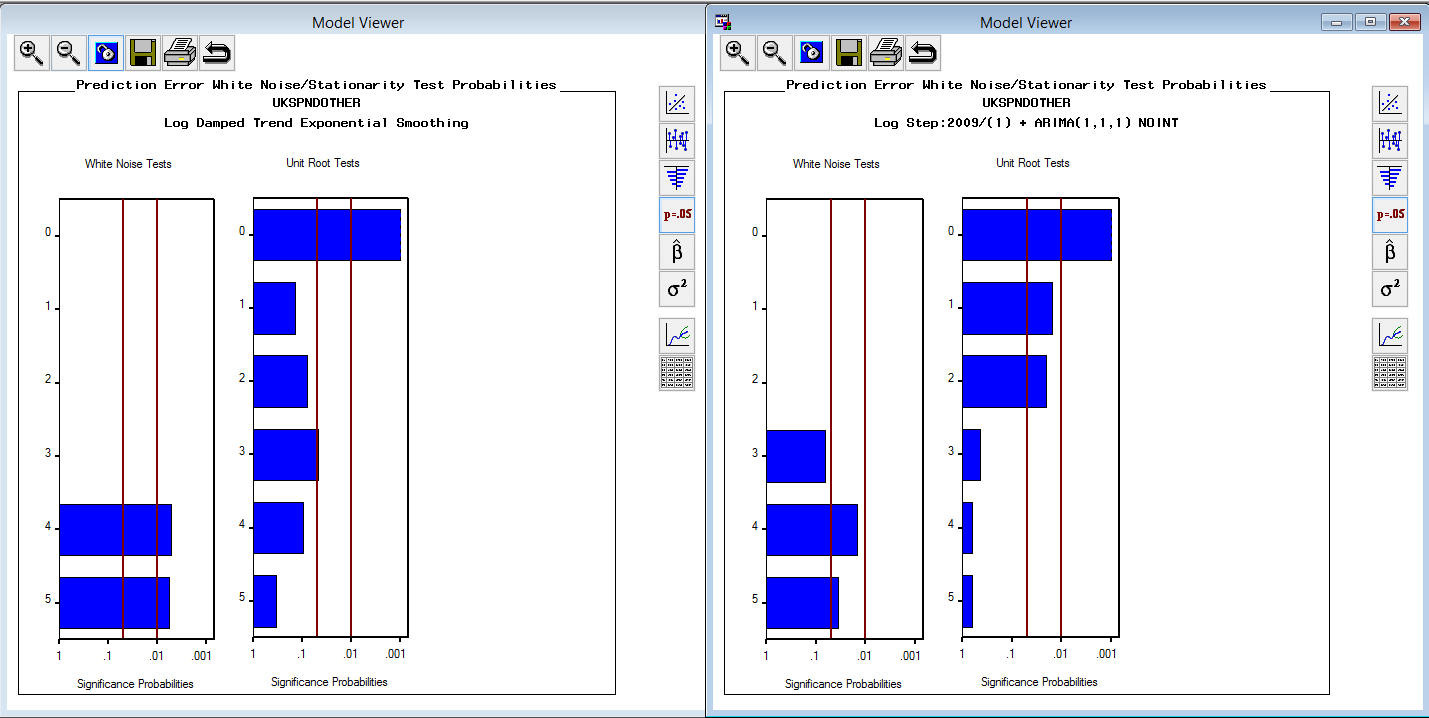
**UK Residents spending in other regions**

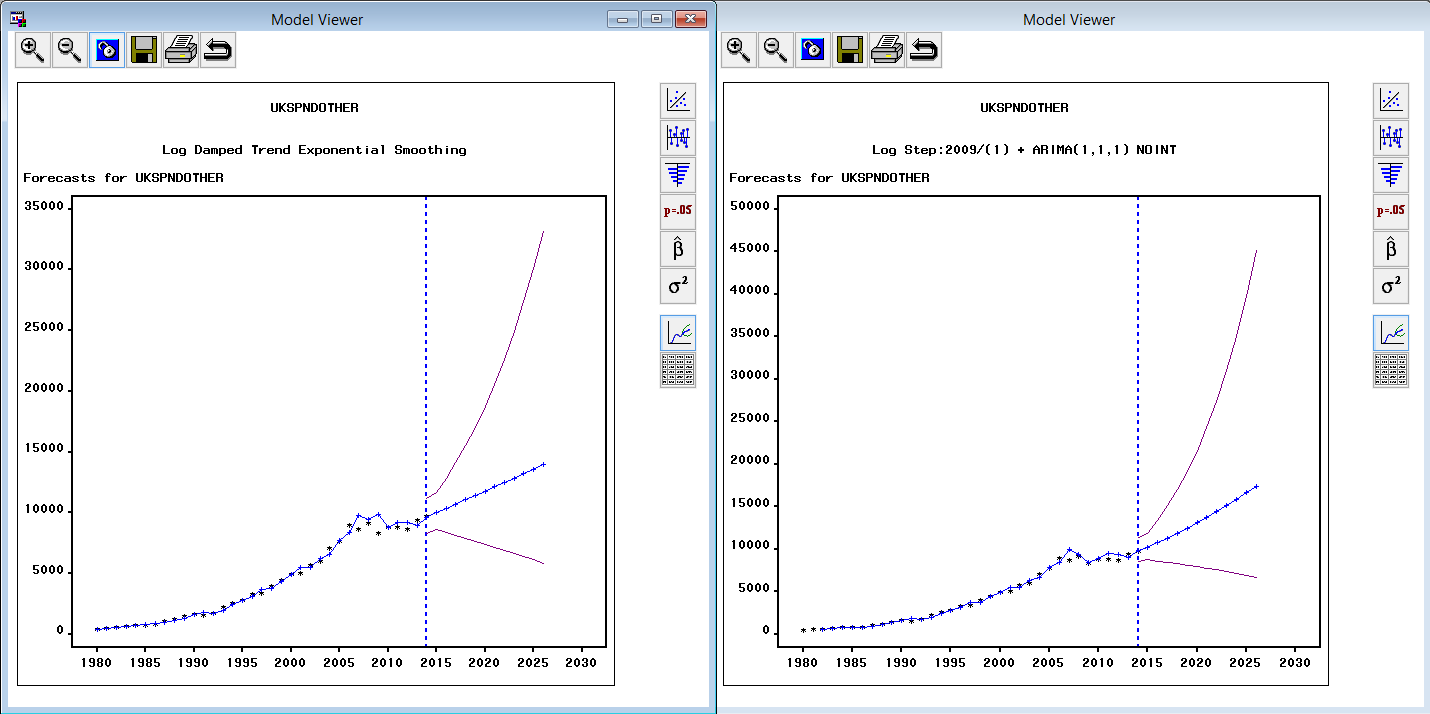
Many model combinations were tried, among which ARIMA (1,1,1) with STEP intervention exponential with LOG transformation is selected as the best fitting model after examining ACF, PACF and IACF plots etc.







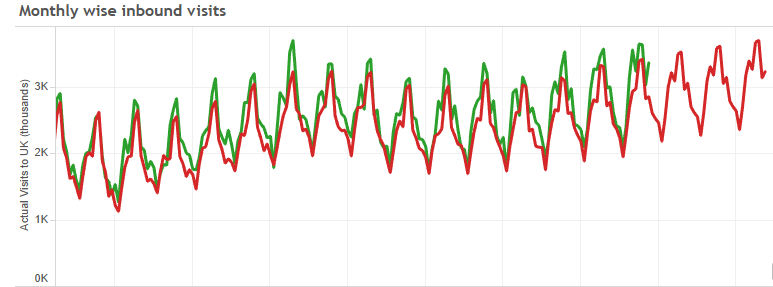




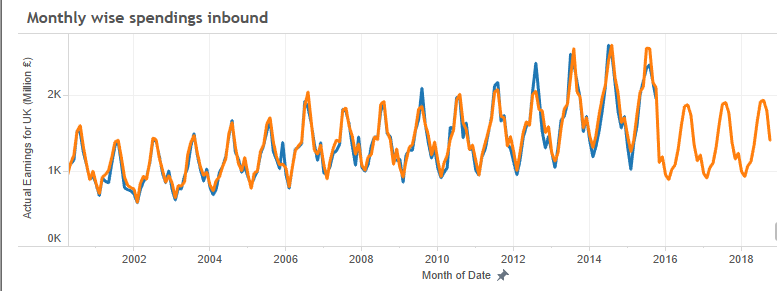
## Visualization Using Tableau

**Monthly Statistics:**

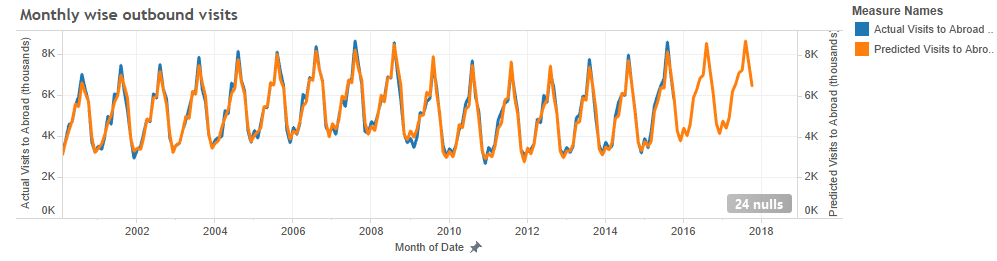
Monthly wise inbound visits- Graph to depict actual vs predicted



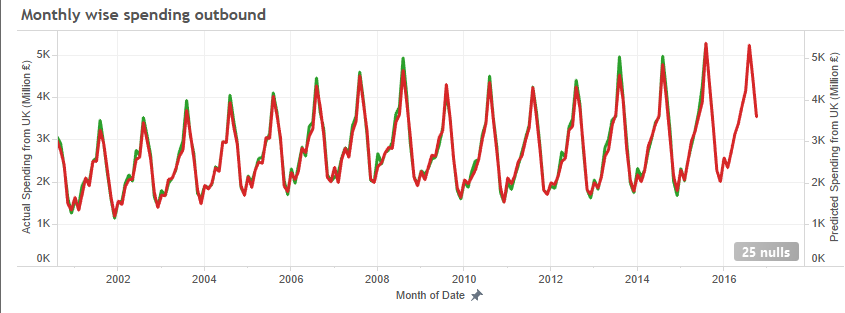
Monthly wise inbound spending’s – Graph to depict actual vs predicted earnings values



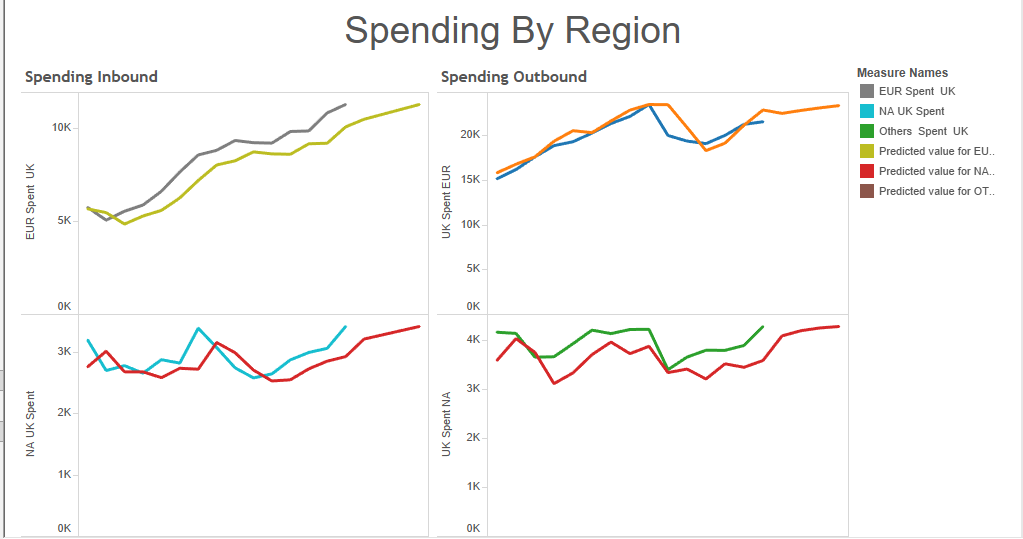
Monthly wise Outbound Visits – Graph to depict actual vs predicted visits to abroad

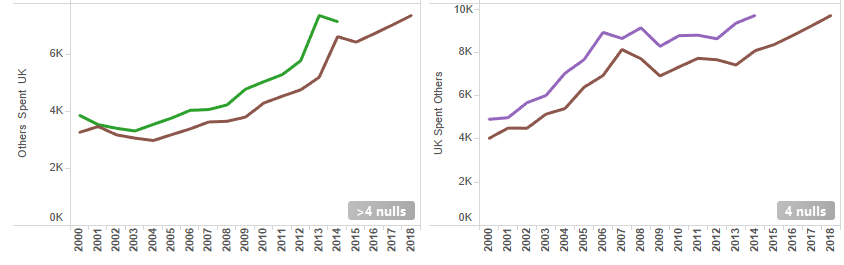


Monthly wise outbound spending



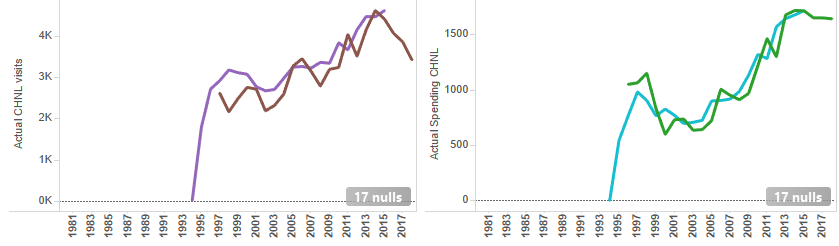
Comparison across inbound and outbound on basis of spending by region

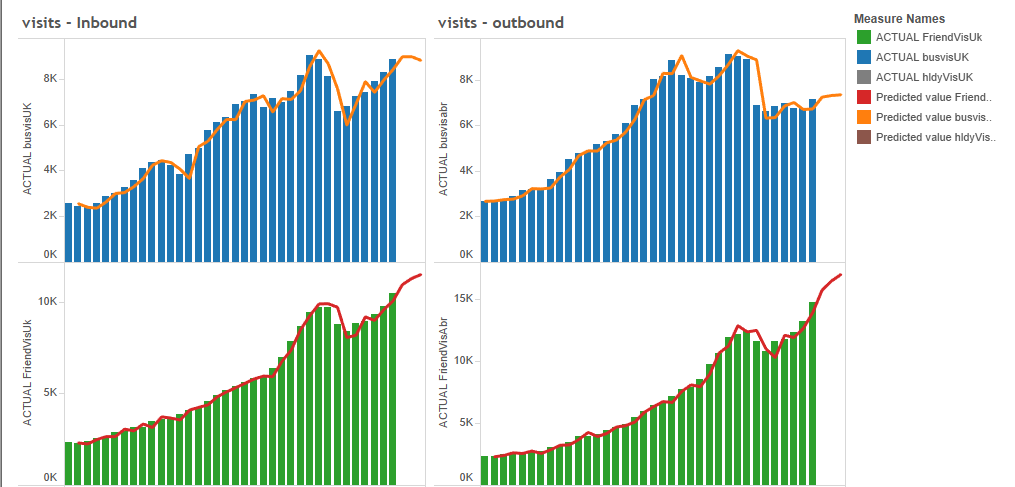


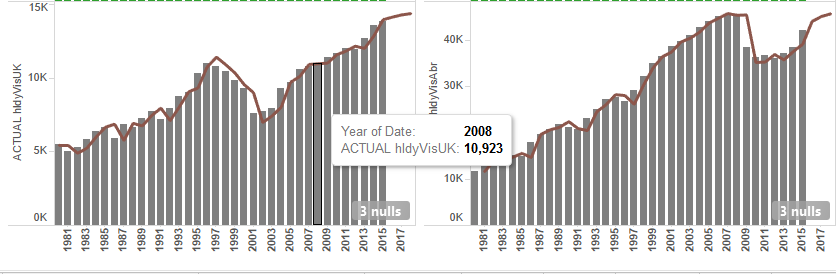


Mode of transport analysis for both inbound visits and spending





Graph to analyze actual visits vs predicted visits by purpose of visits



SAS Codes & Data Sets:  
Import csv files into SAS data files:



Cross correlation SAS code:



Initial and Final Data Sets:

