Retail Giant Sales Forecasting using Time Series Analysis – A Case Study

Group Name:
1.Thota Siva Praveen
2.Ujwal Kiran Bhargava
3.Vijayaraghavan
4.Bhavani Bhavineni

Introduction, Background and Problem Statement

- “Global Mart” is an online store super giant having worldwide operations. It takes orders and delivers across the globe and deals with all the major product categories - consumer, corporate & home office.
- The main objective is to forecast the sales and the demand for the next 6 months using time series analysis in order to manage revenue and inventory accordingly.
- The store caters to 7 different market segments and in 3 major categories.(i.e. 21 segments in total)
- This is done by finding out the 2 most profitable (and consistent) segments from the 21 total segments and forecasting the sales and demand for these segments.
- For this purpose, the coefficient of variation of the profit for all 21 market segments is used as the metric.

Problem Solving Methodology

Business Understanding
(understand the different consumer segments and regions and find out the two most profitable segments)

Data Understanding and Preparation
(segment dataset into 21 segments based on the 7 markets and the 3 consumer segment levels and convert this data into time series)

Forecasting
(Find out the two most profitable segments using coefficient of variation of profit and forecast sales and quantity for the next 6 months)

Model Building
(Build the resulting model using classical decomposition and auto ARIMA after smoothening the data)

Model Evaluation
(Forecast the sales and quantity for next 6 months using the model obtained and evaluate the same using MAPE)

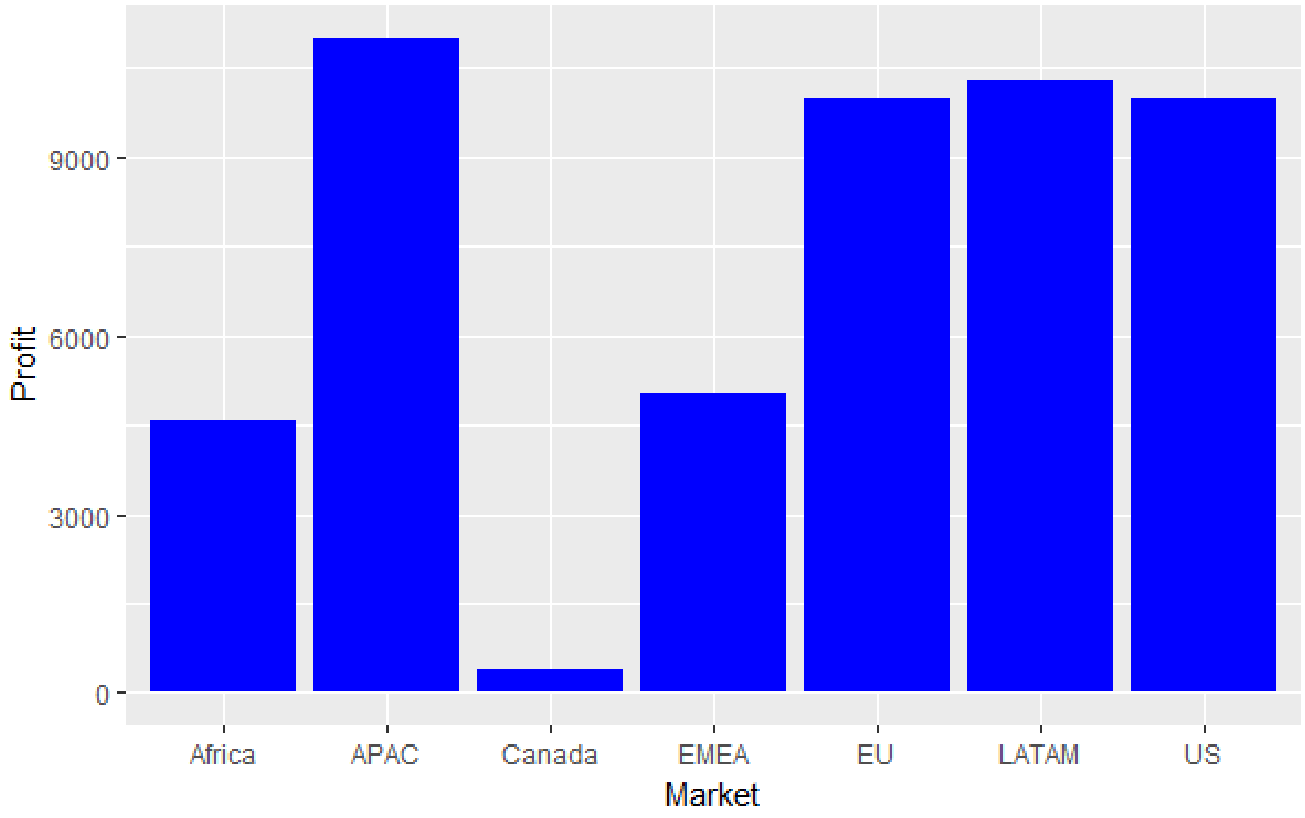
Understanding Data and Assumptions Used in this Case Study

- The given .csv file used in this case study represents transaction level data, where each row represents a particular order made on the online store.
- There are 24 attributes related to each such transaction. The “Market” attribute has 7-factor levels representing the geographical market sector that the customer belongs to. The “Segment” attribute tells which of the 3 segments that customer belongs to.
- Thus, there are 21 overall segments based on market and customer segment types.
- However, only the best two profitable (and consistent) segments are selected using coefficient of variation of profit and used for forecasting and analysis.
- Coefficient of variation of profit is the standard deviation of the profit of the segments divided by mean of all the profits.

Understanding Data and Assumptions Used in this Case Study

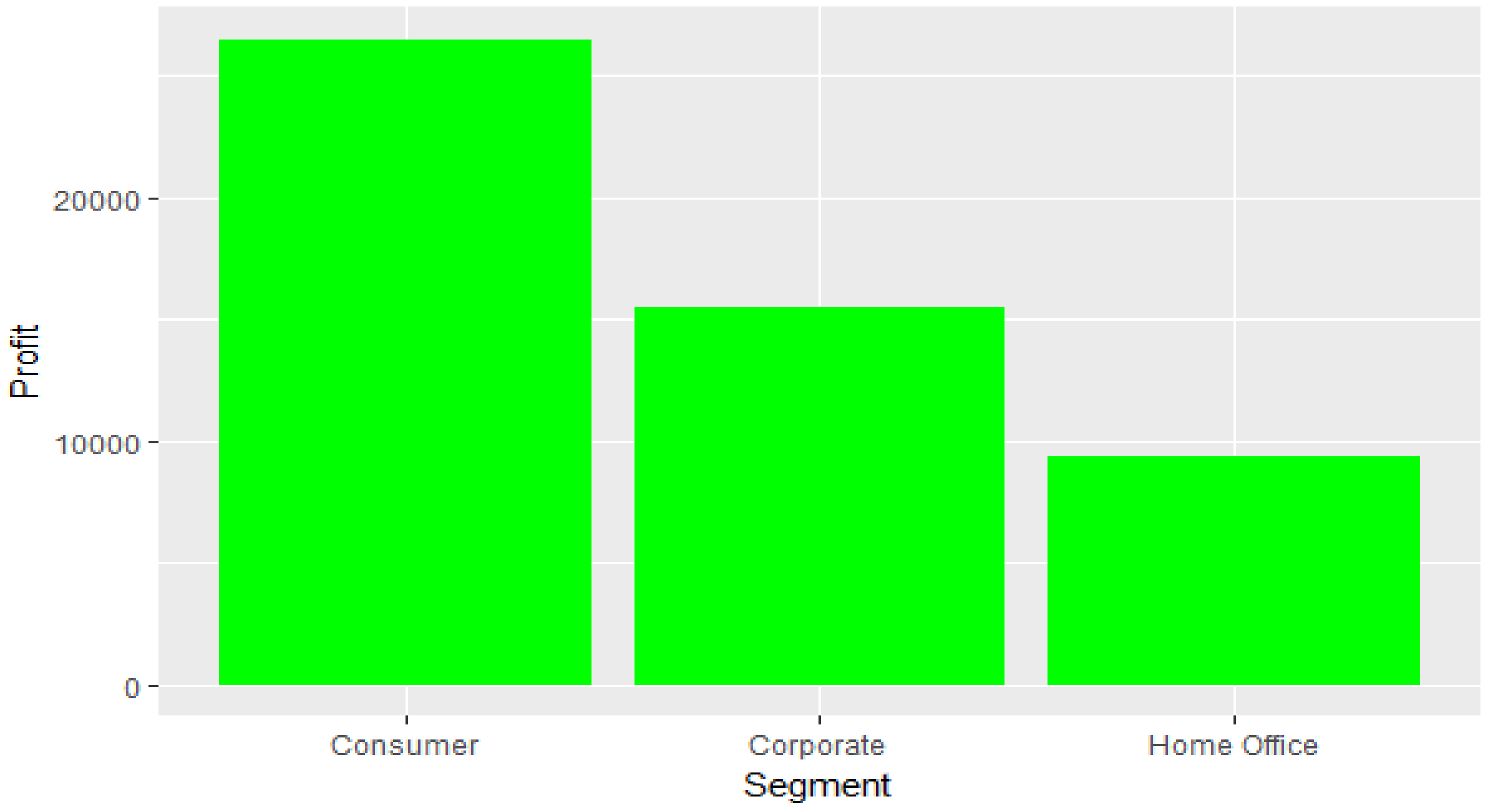
- The columns having all NA’s or 0’s as row value are discarded.
- Row.ID column is discarded as it is not relevant for the analysis.
- Postal.Code column is discarded as it has most of the NA values.
- All required character type variables are converted into factor type for ease of analysis.
- All month-year type variables are converted into appropriate date format for ease of analysis.
- The time series are modelled only after converting them into stationary ones.

Exploratory Data Analysis



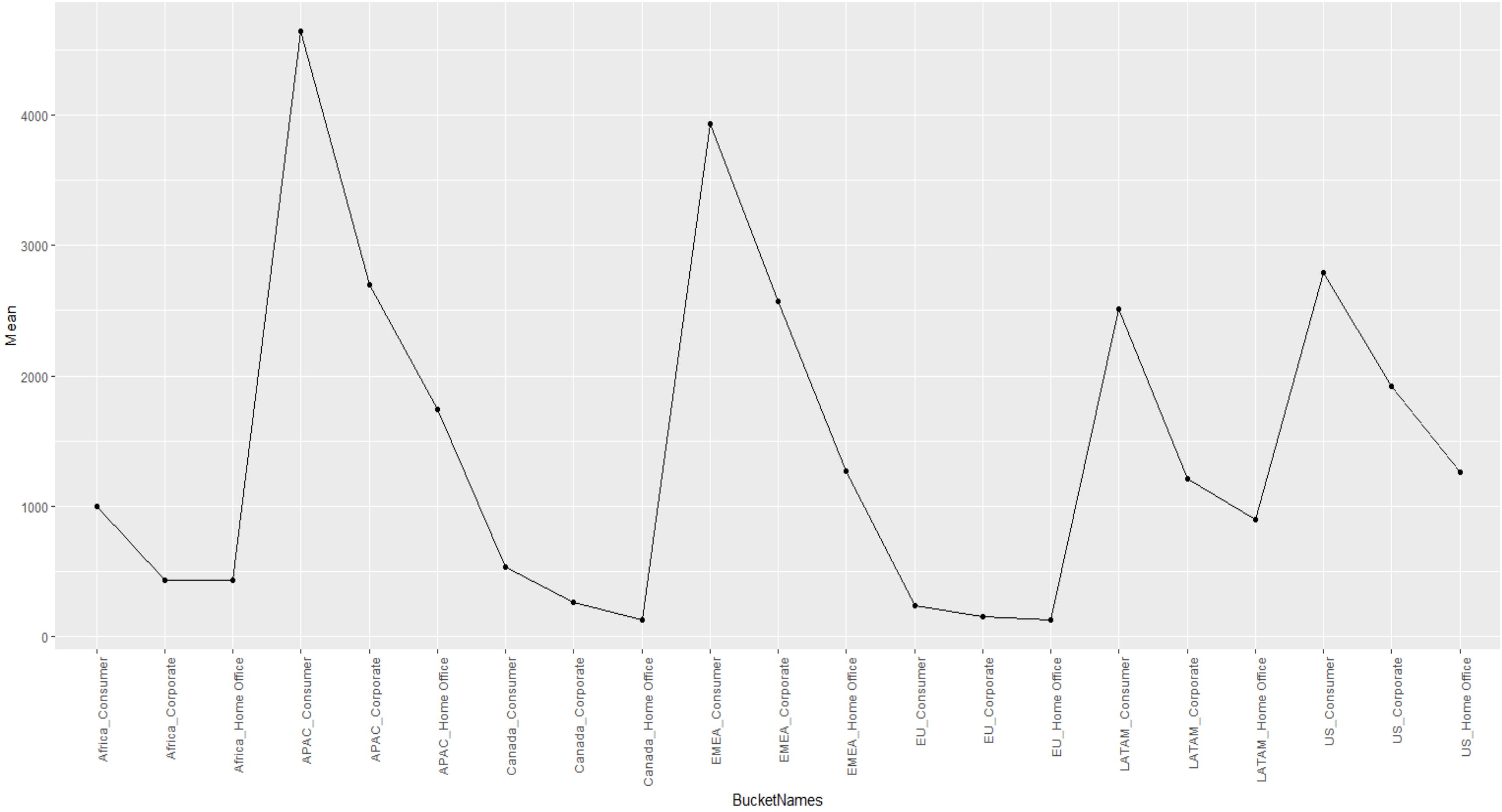
From the graph, it is clear that the APAC geographical market region is the most profitable.

Exploratory Data Analysis



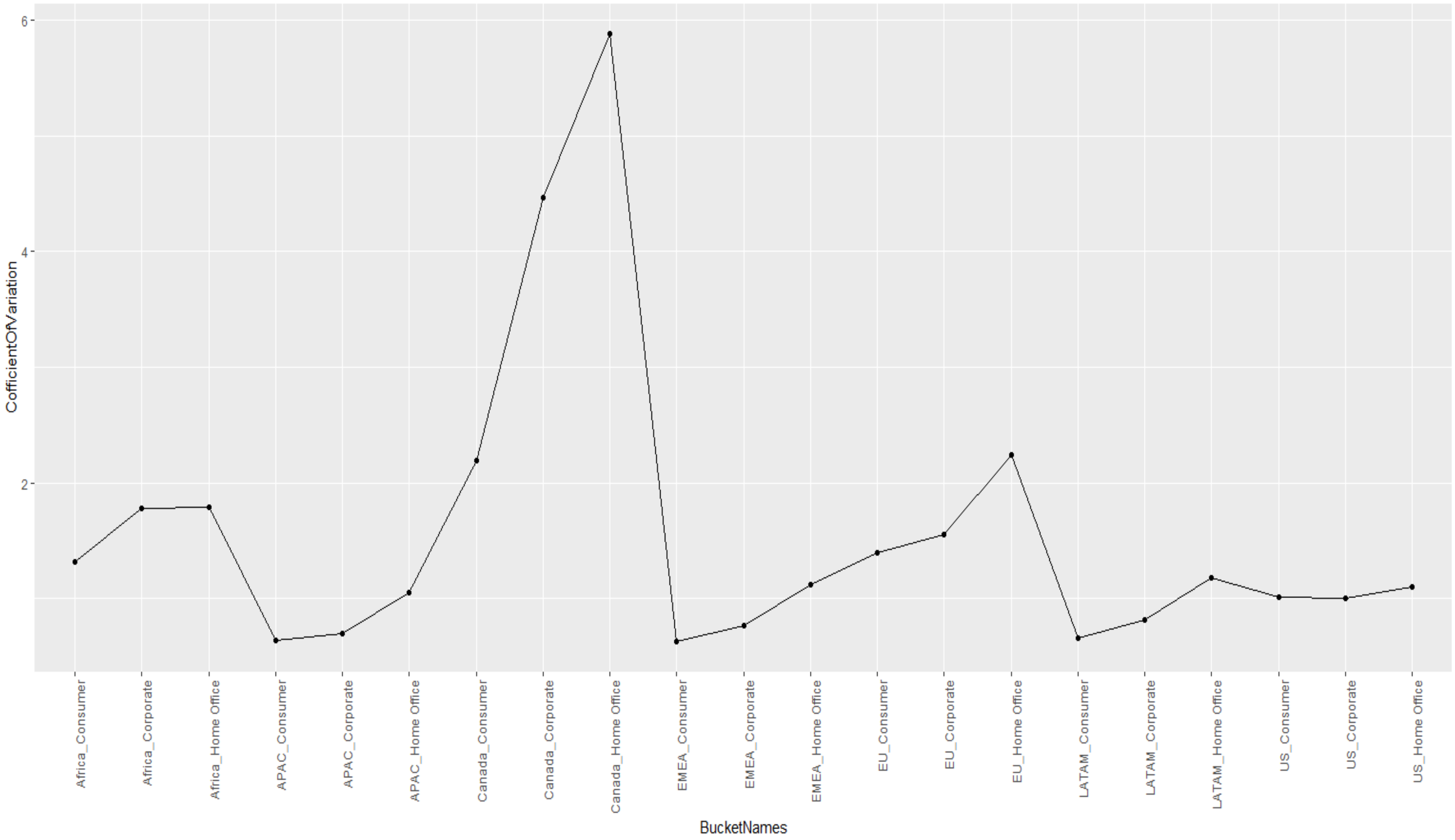
From the graph, it is clear that the Consumer segment is the most profitable.

Exploratory Data Analysis



From the graph, it is clear that the customer segment in the APAC region has the highest average profit (mean).

Exploratory Data Analysis



From the graph, it is clear that the customer segment in the APAC region has the lowest coefficient of variance (mean).

Selecting the most profitable and consistent segments

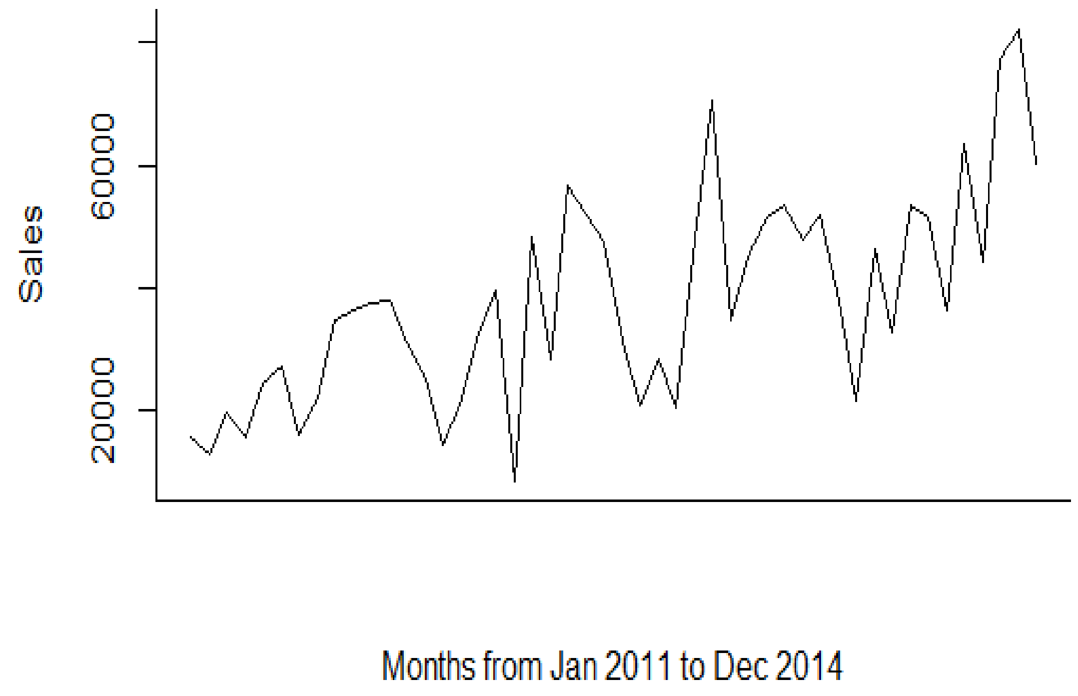
- From our observations, it is clear that APAC consumer and EU consumer segments are the most profitable and consistent ones out of the 21 segments.
- This is because, APAC consumer and EU consumer segments have the lowest coefficient of variation of profit (0.63 and 0.62 respectively).

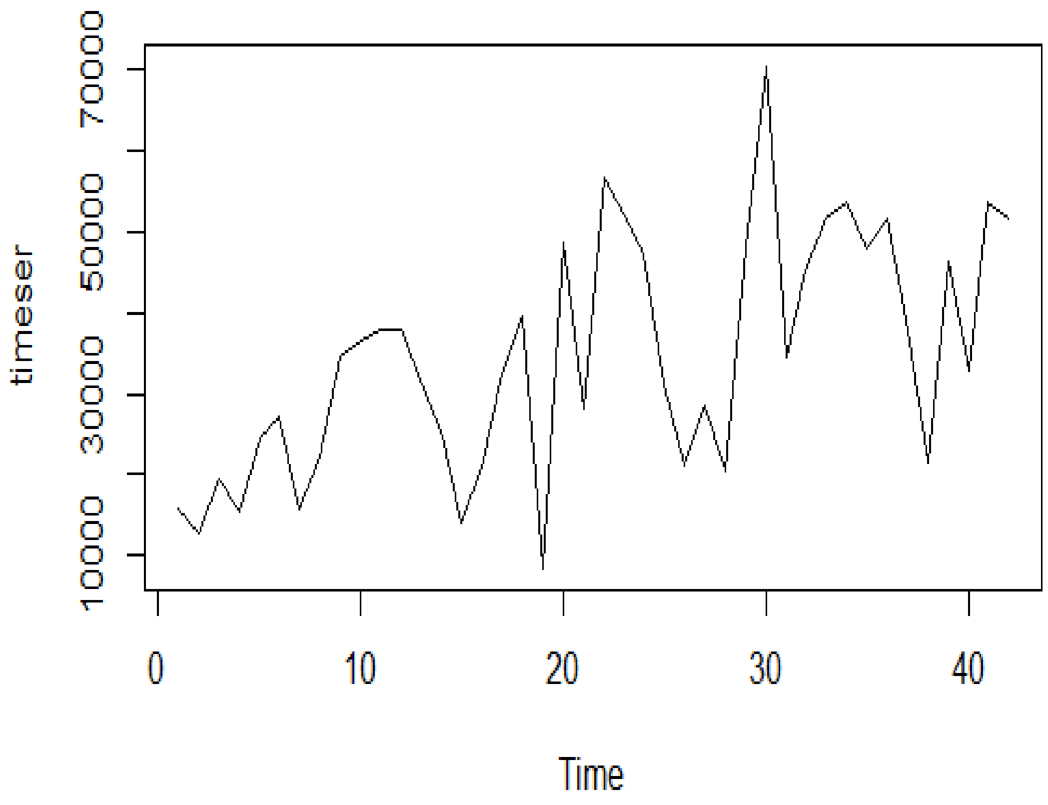
Model Building and Evaluation

The following steps are involved in model building and evaluation:

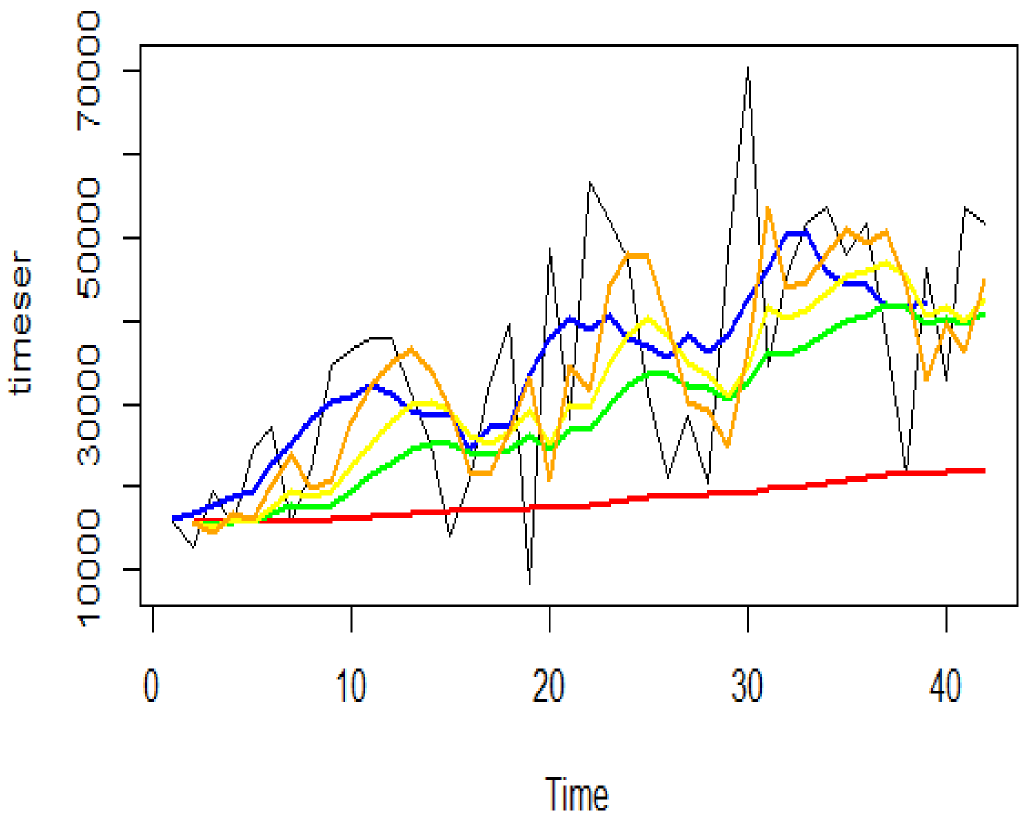
- Convert transaction-level data into time series.
- Smoothen the time series using exponential and classical decomposition methods.
- Model trend and seasonality.
- Model the stationary component using ARMA.
- Analyse the diagnostic plots, ACF and PACF.
- Plot the predicted values and forecast the ame using ARIMA
- Evaluate the model using MAPE.

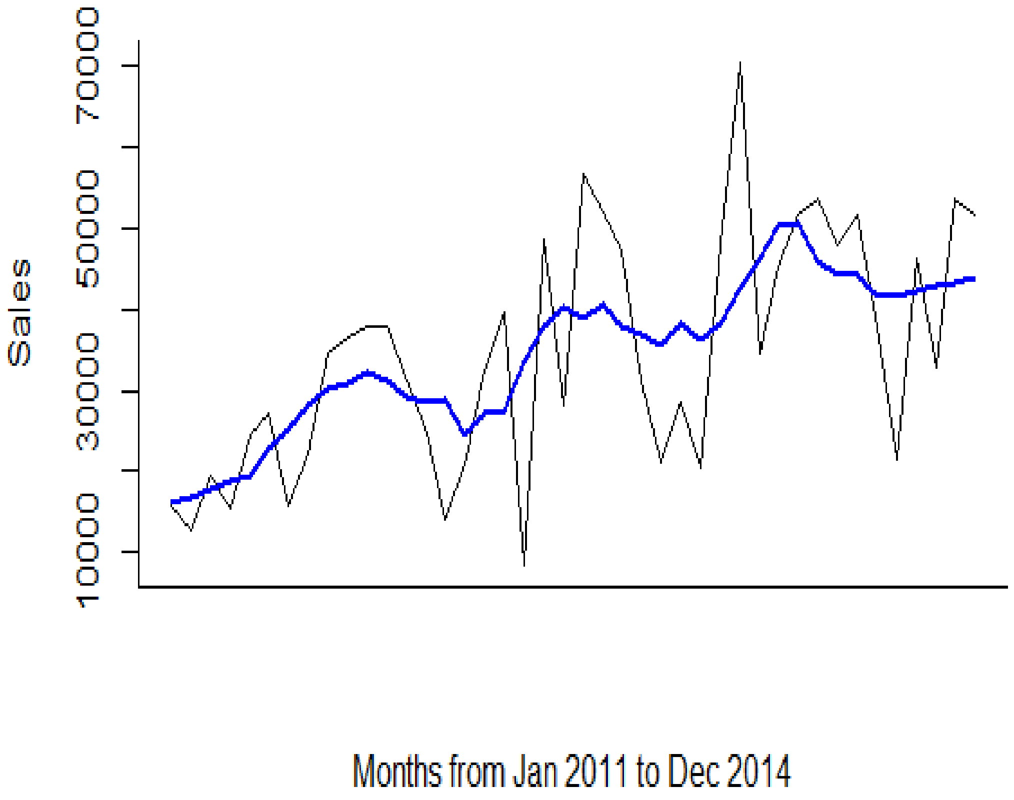
APAC Consumer Sales Time Series



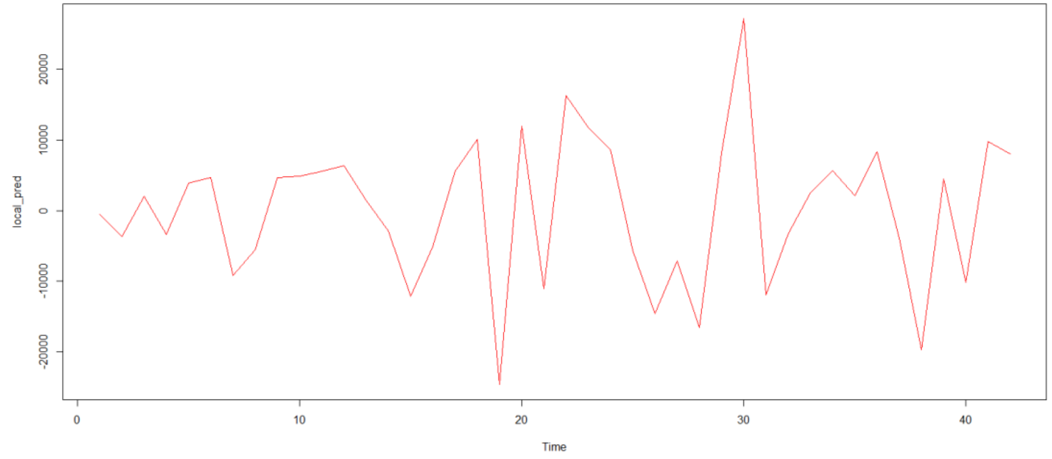


Smoothening for APAC Consumer Sales Time Series

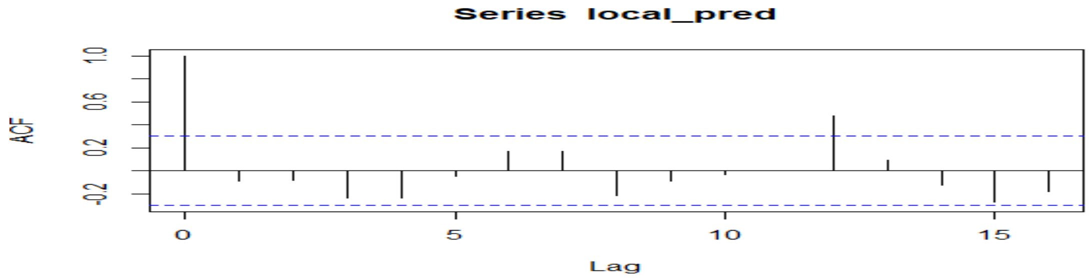


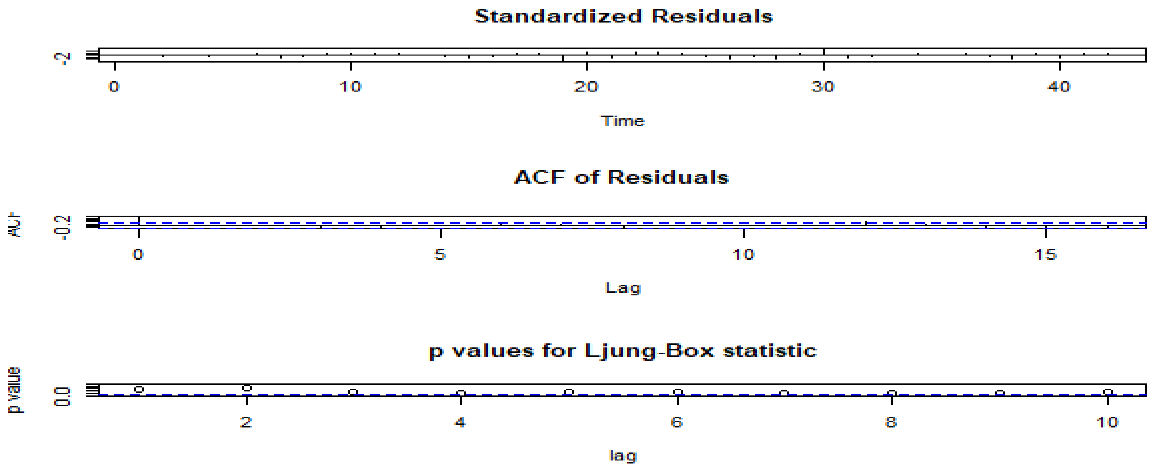


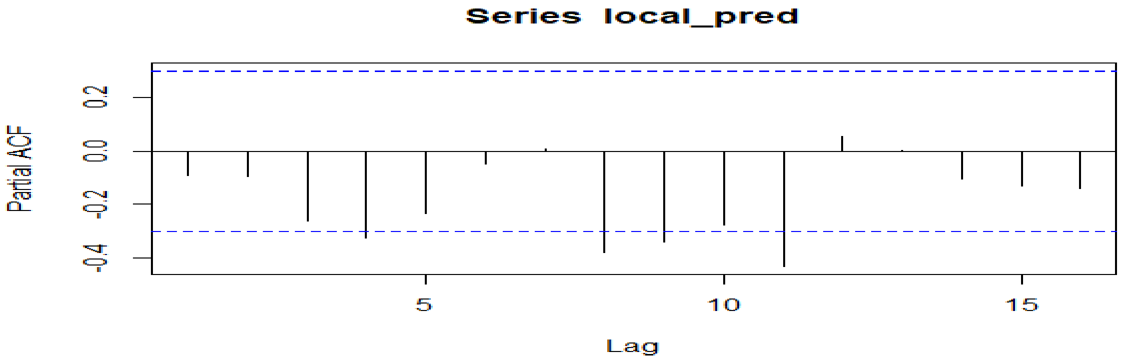
ARMA for APAC Consumer Sales Time Series



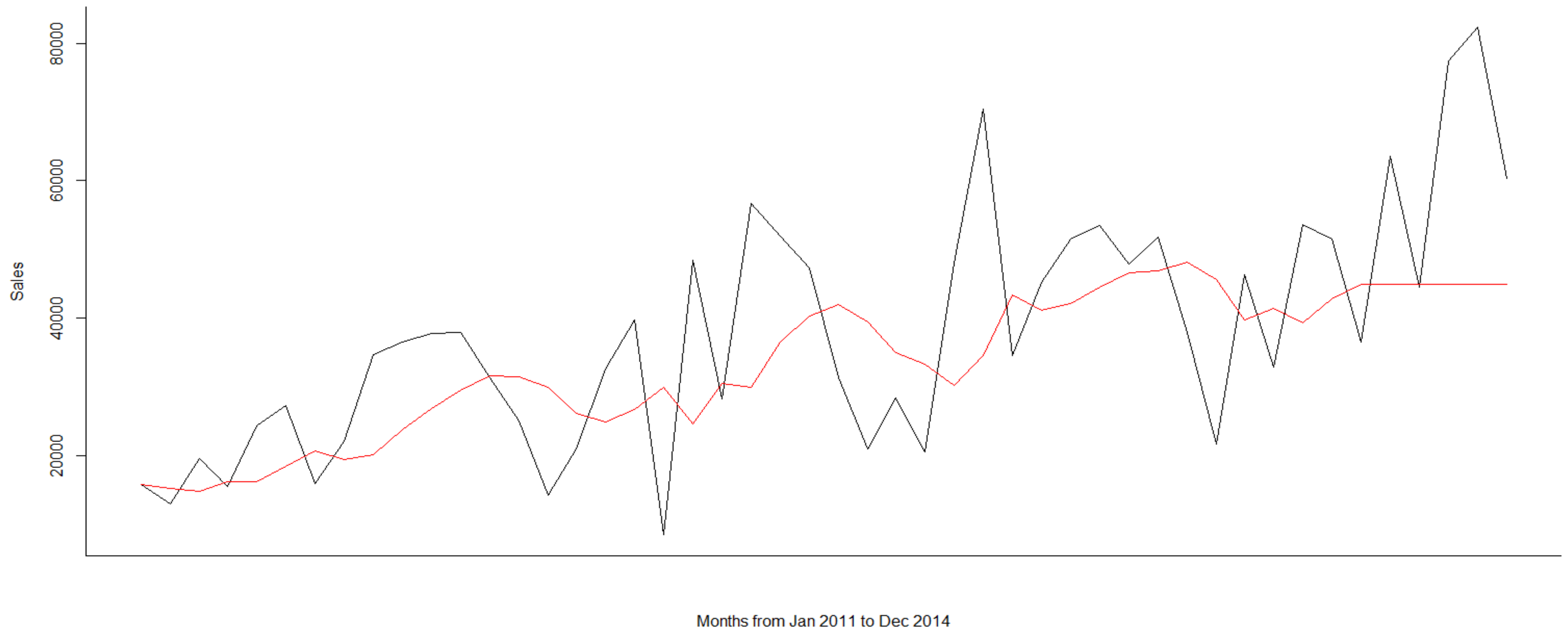
Series: local_pred ARIMA(0,0,0)
log likelihood=-446.83 AIC=895.66
AICc=895.76 BIC=897.4



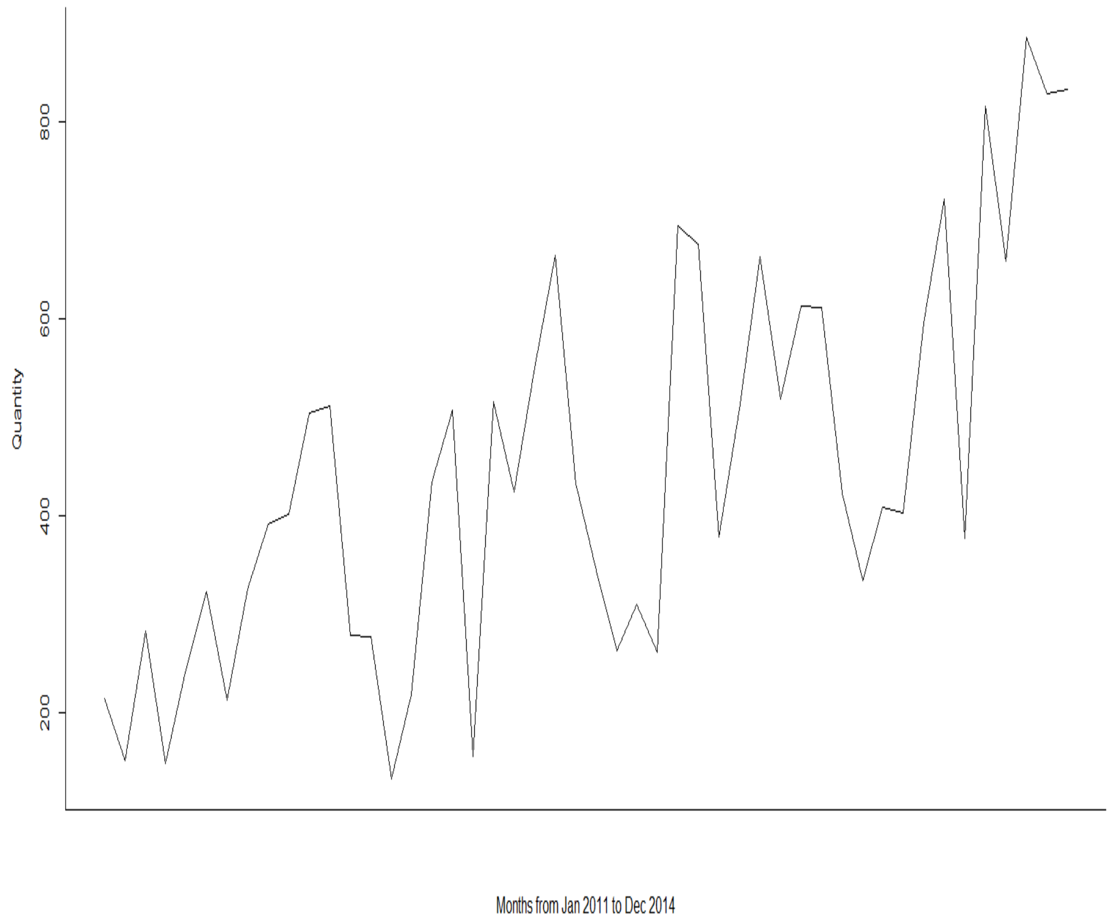


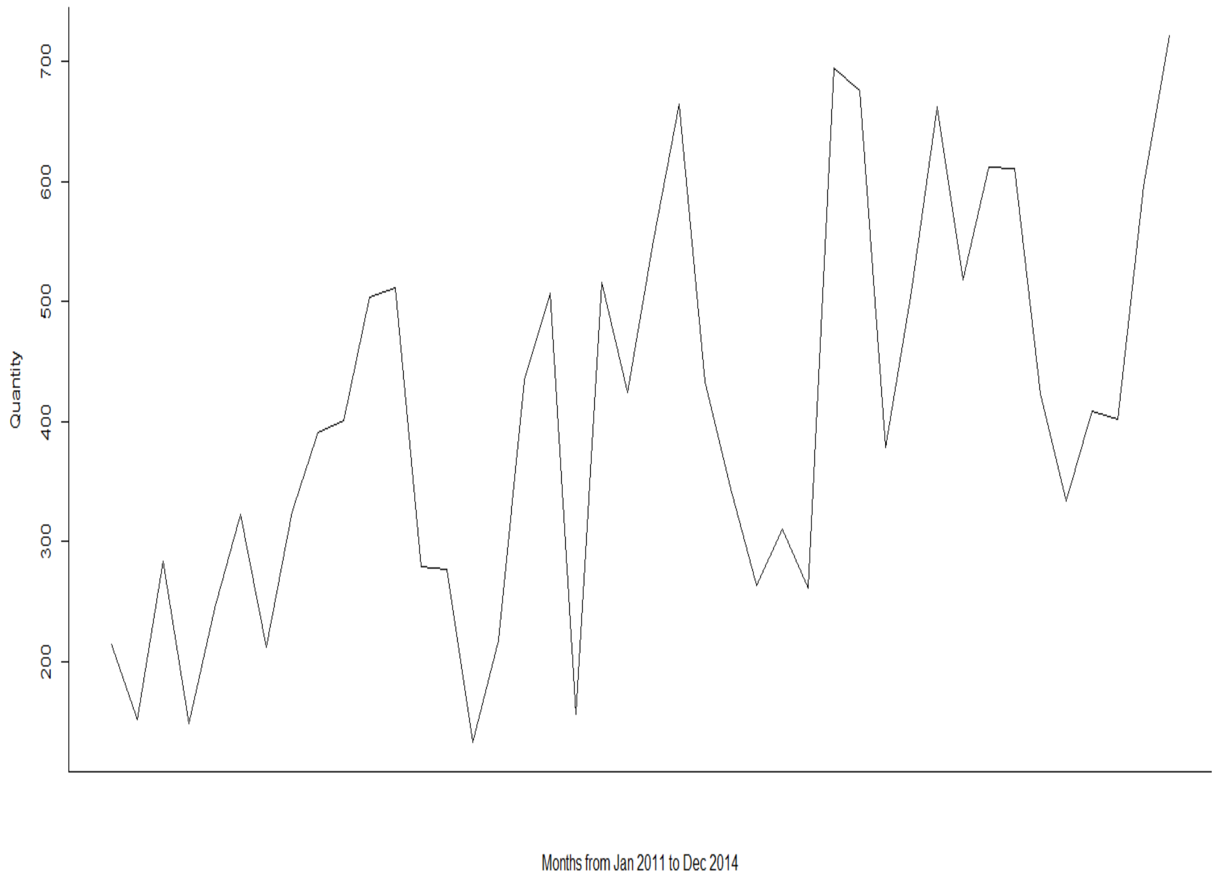


Forecasting APAC Consumer Sales using ARIMA

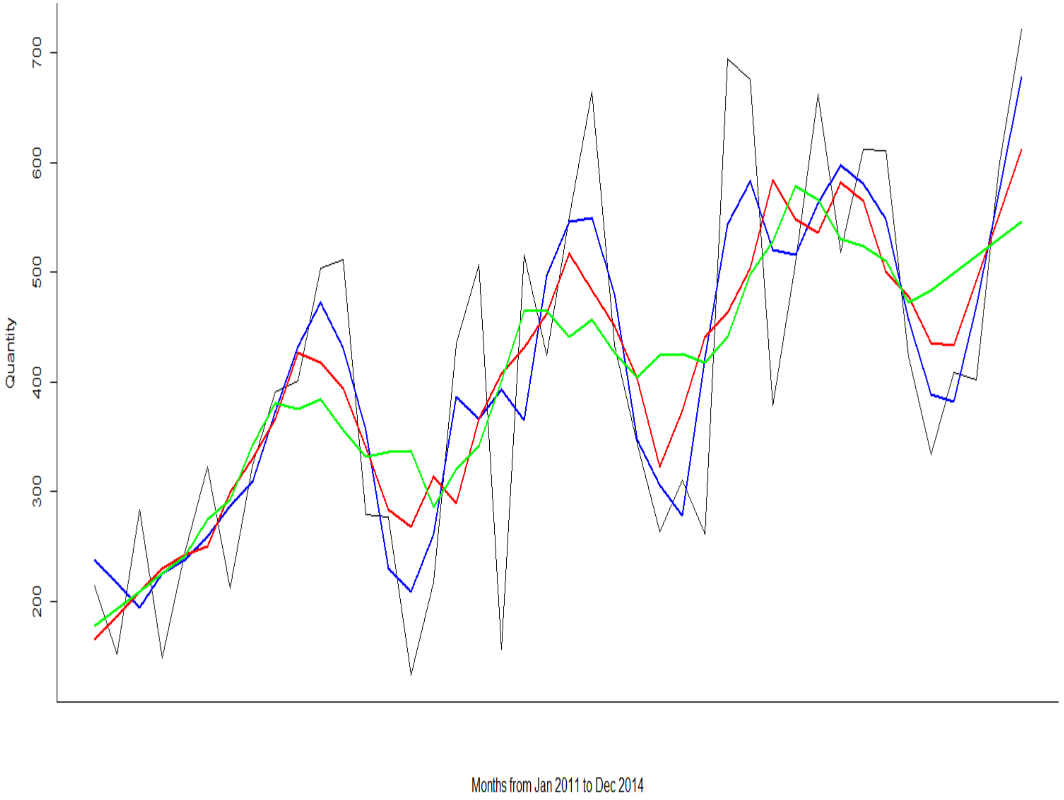


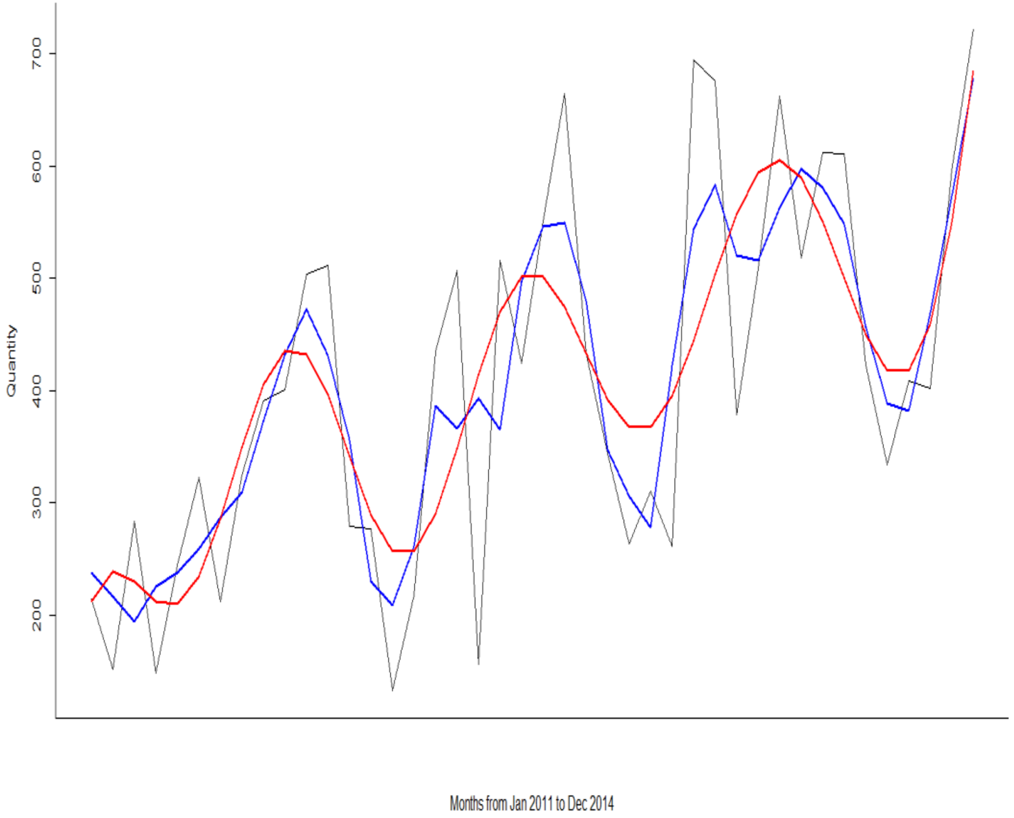
APAC Consumer Quantity Time Series



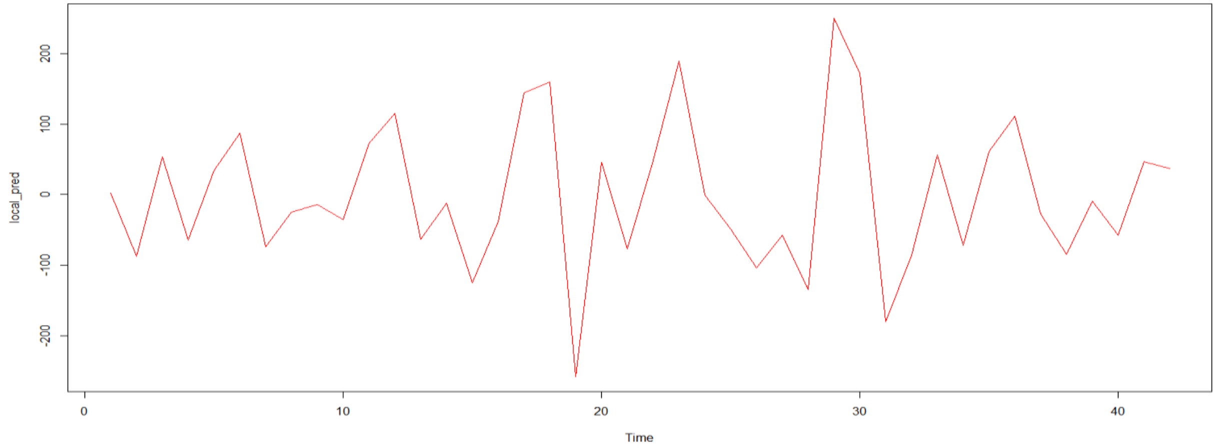


Smoothening for APAC Consumer Quantity Time Series

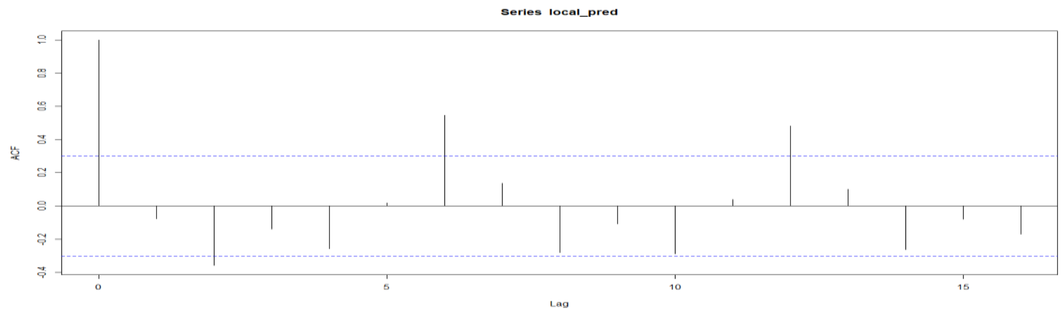


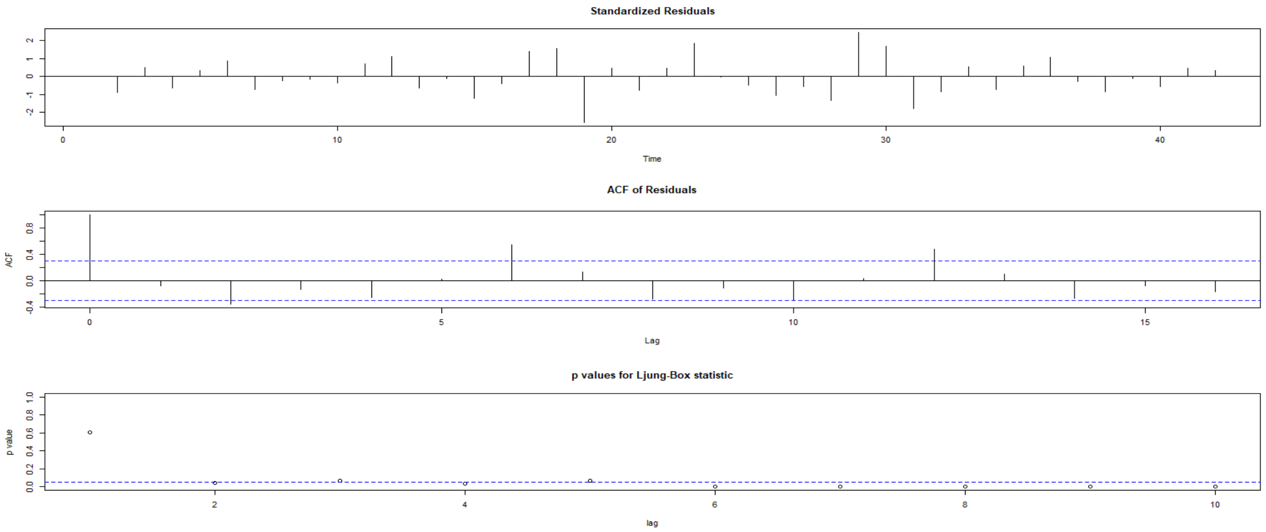


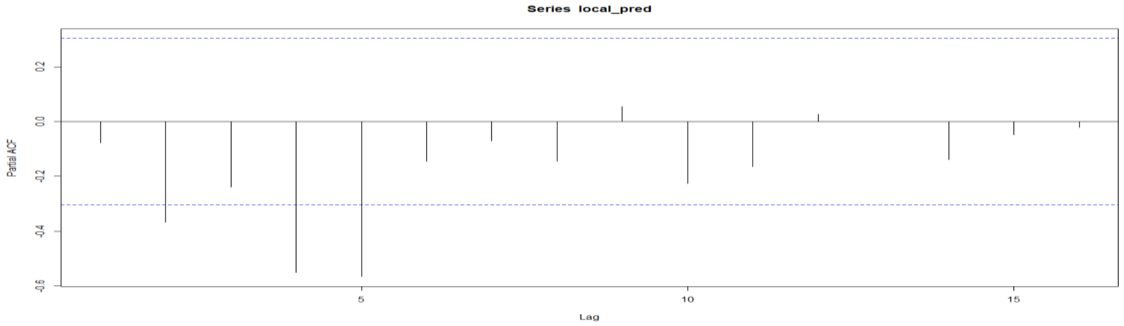
ARMA for APAC Consumer Quantity Time Series



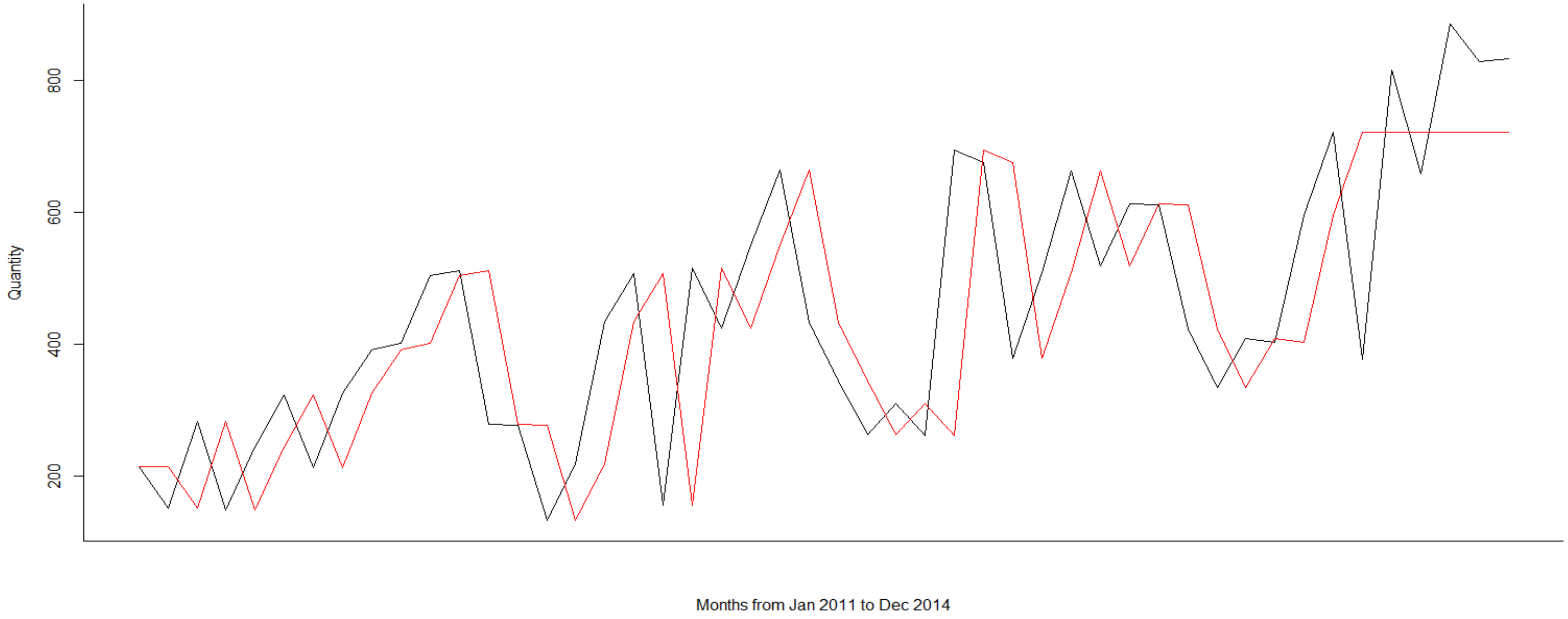
Series: local_pred ARIMA(0,0,0)
log likelihood=-253.71 AIC=509.42
AICc=509.52   BIC=511.16



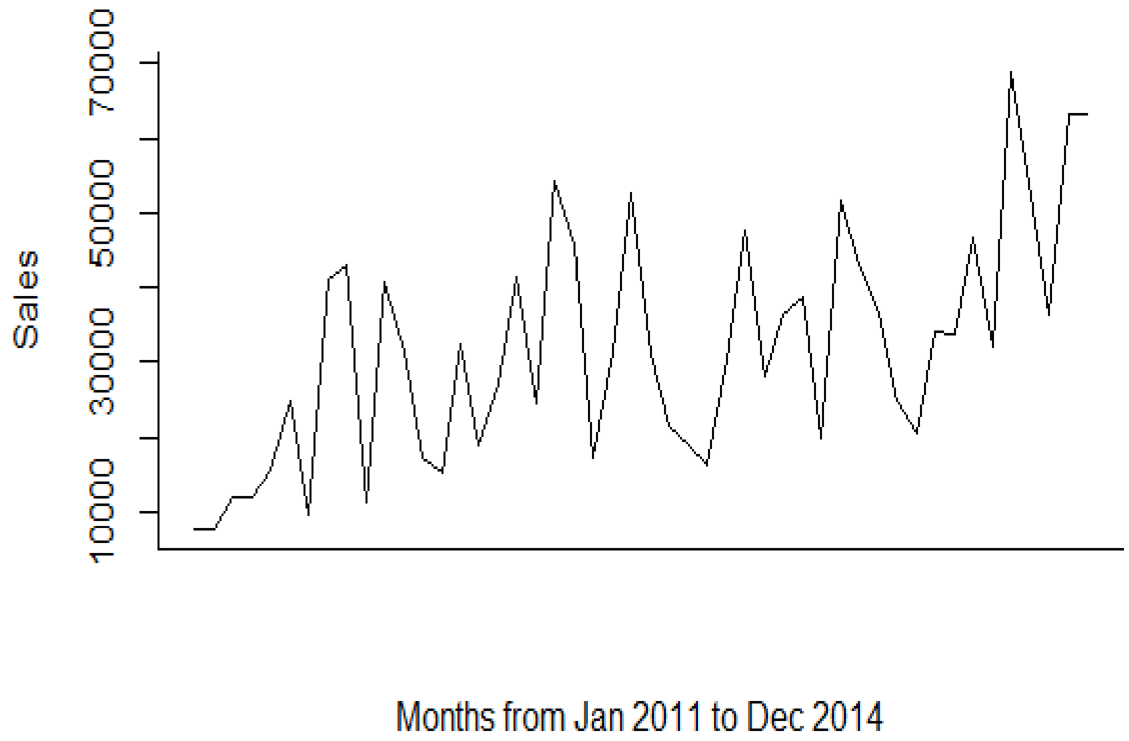


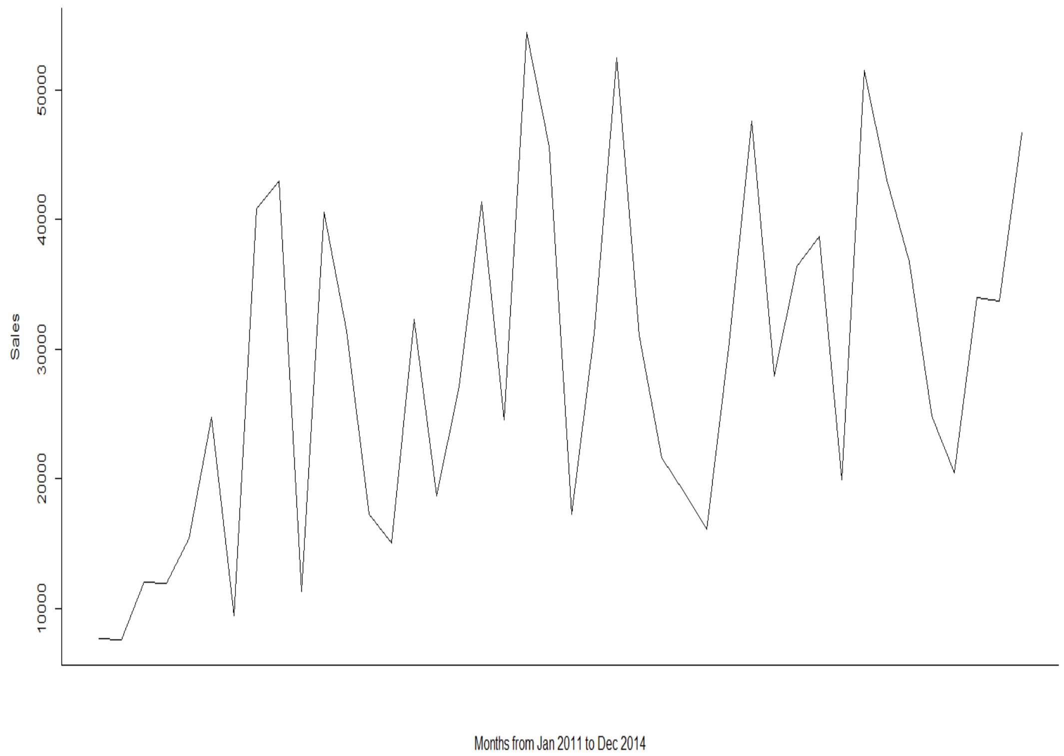


Forecasting APAC Consumer Quantity using ARIMA

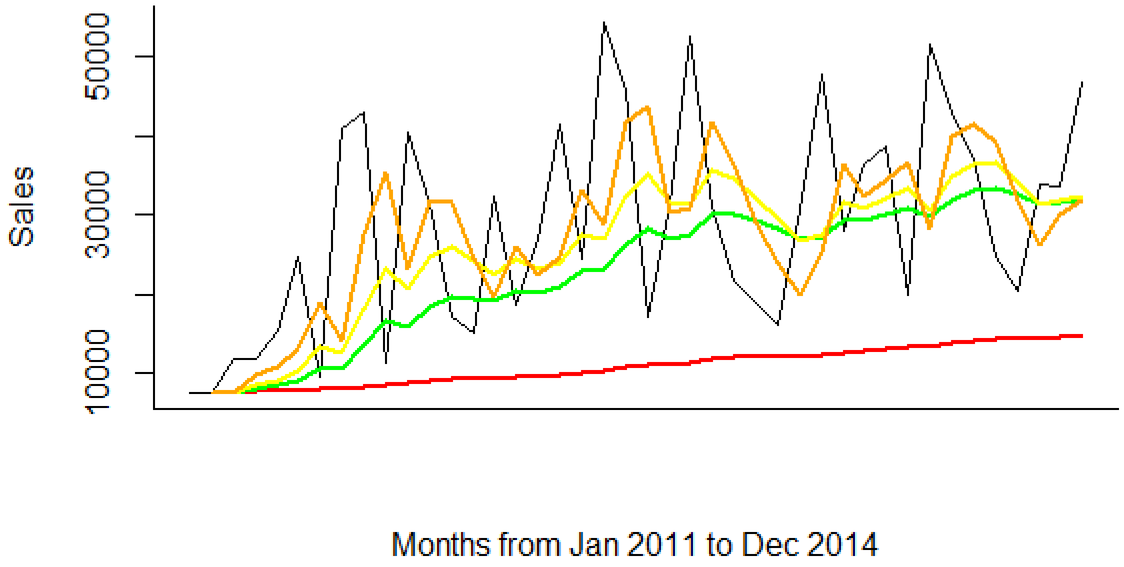


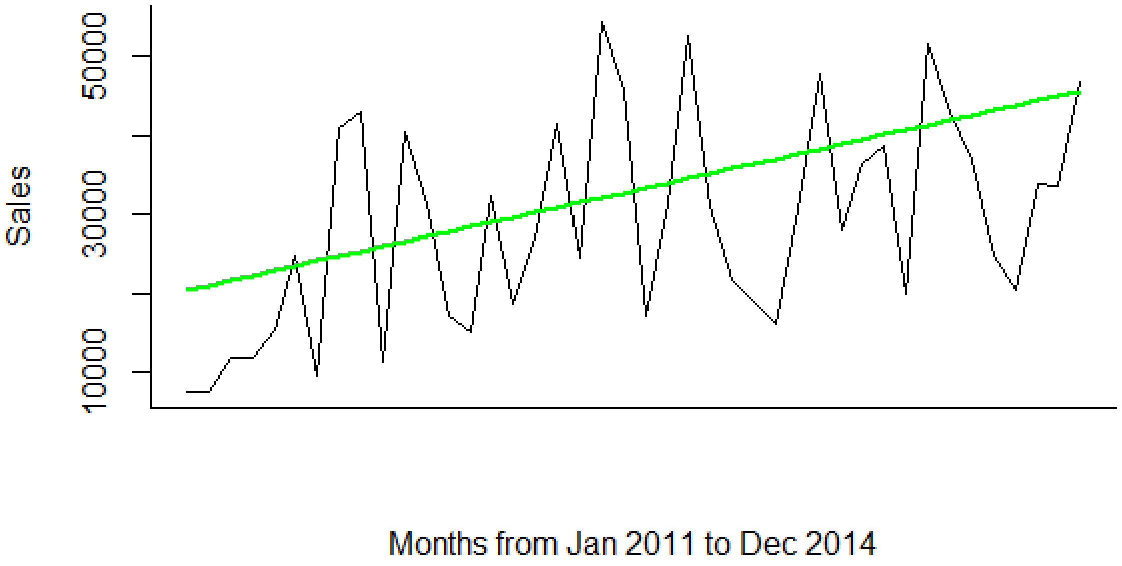
EU Consumer Sales Time Series



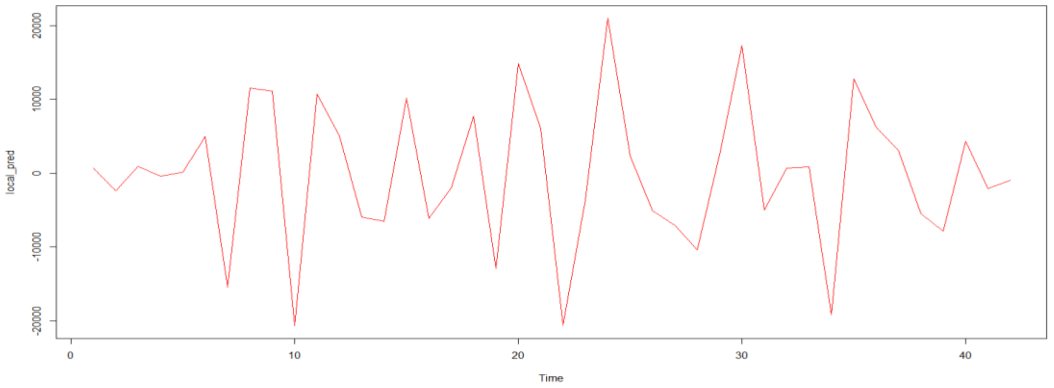


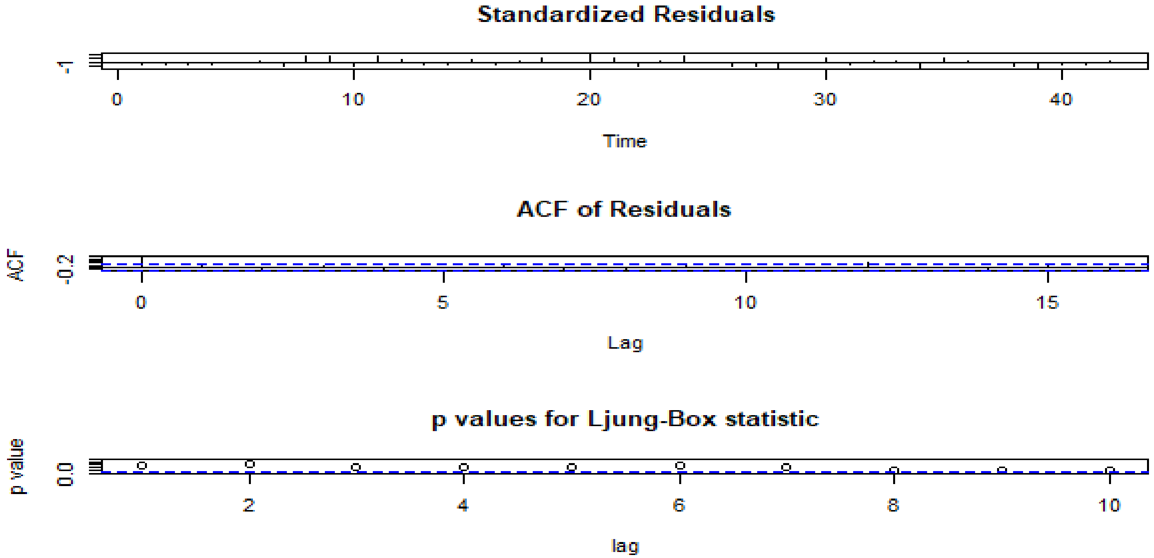
Smoothening for EU Consumer Sales Time Series



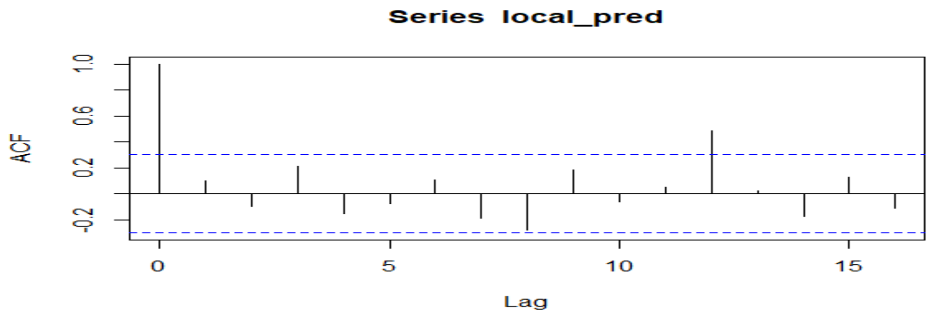


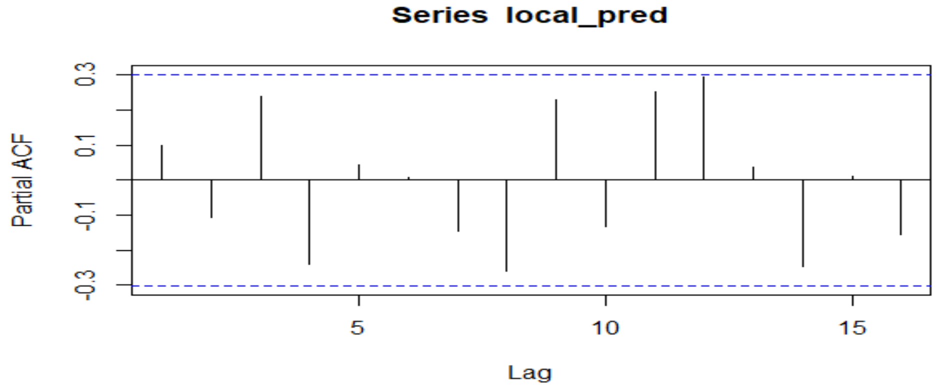
ARMA for EU Consumer Sales Time Series



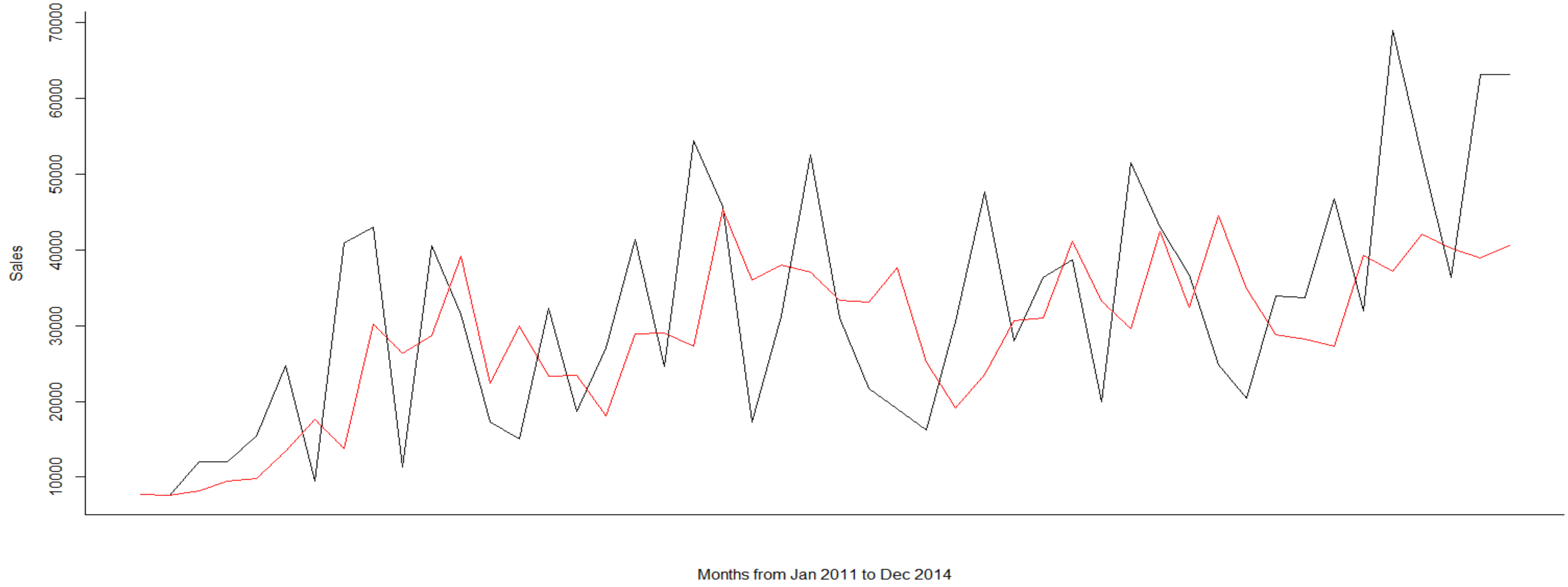


Series: local_pred ARIMA(0,0,0)
log likelihood=-453.37
AIC=910.75 AICc=911.06
BIC=914.23

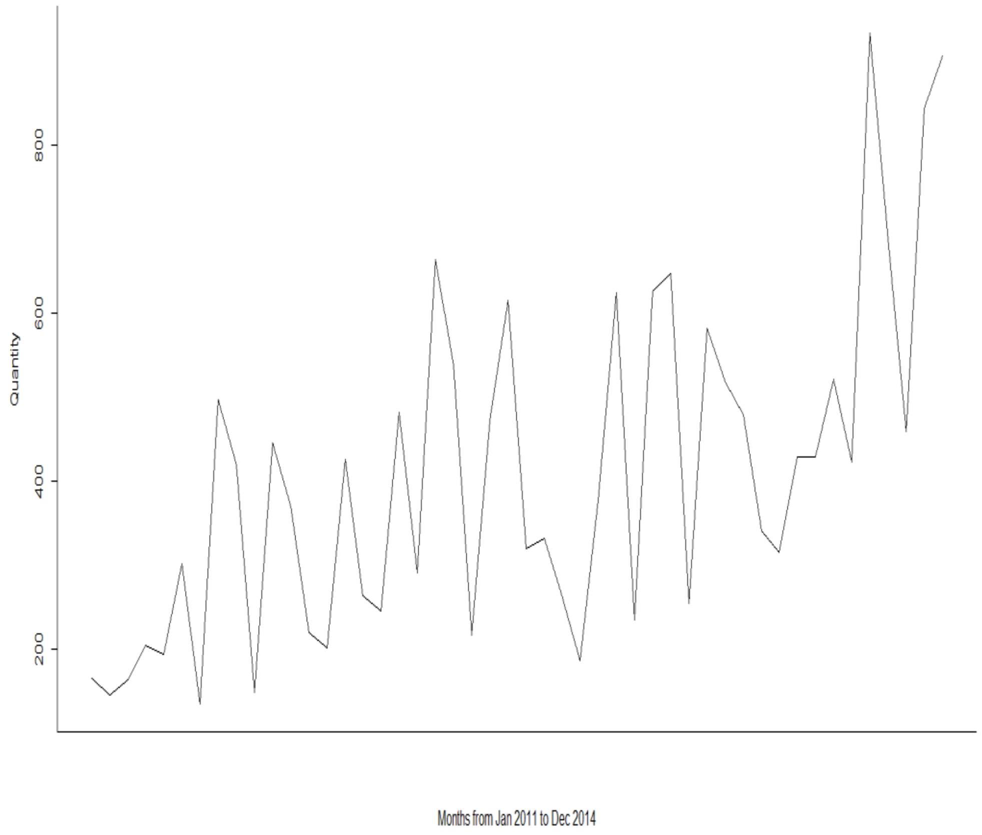


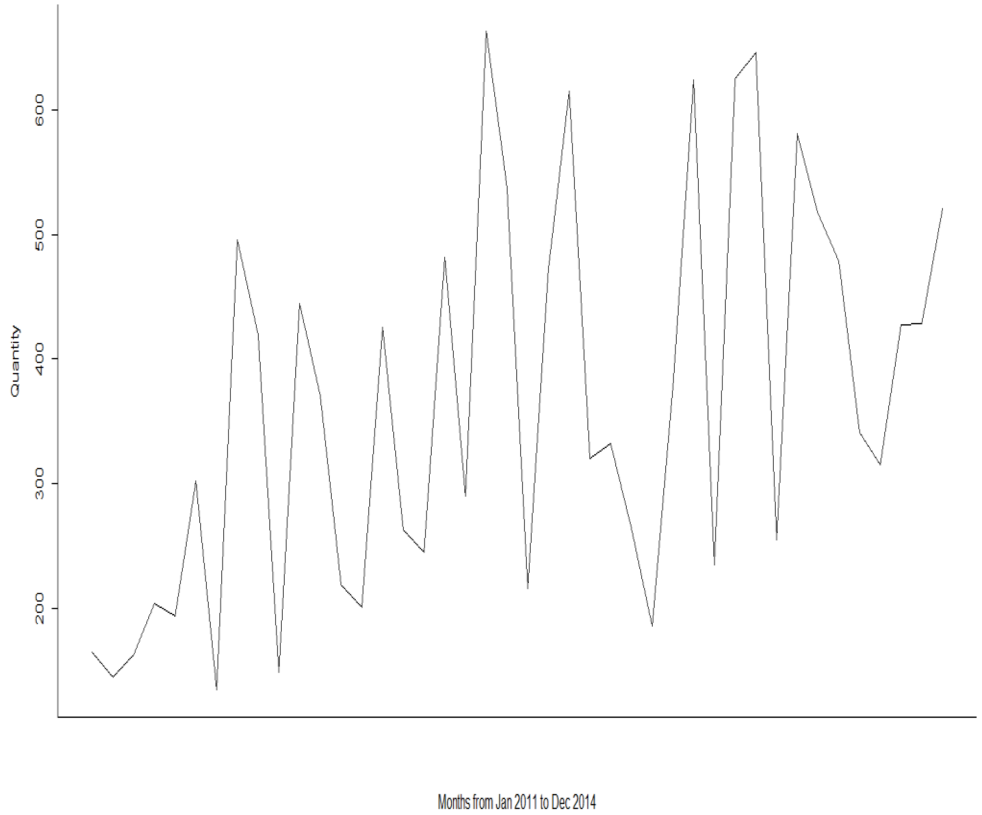


Forecasting EU Consumer Sales using ARIMA

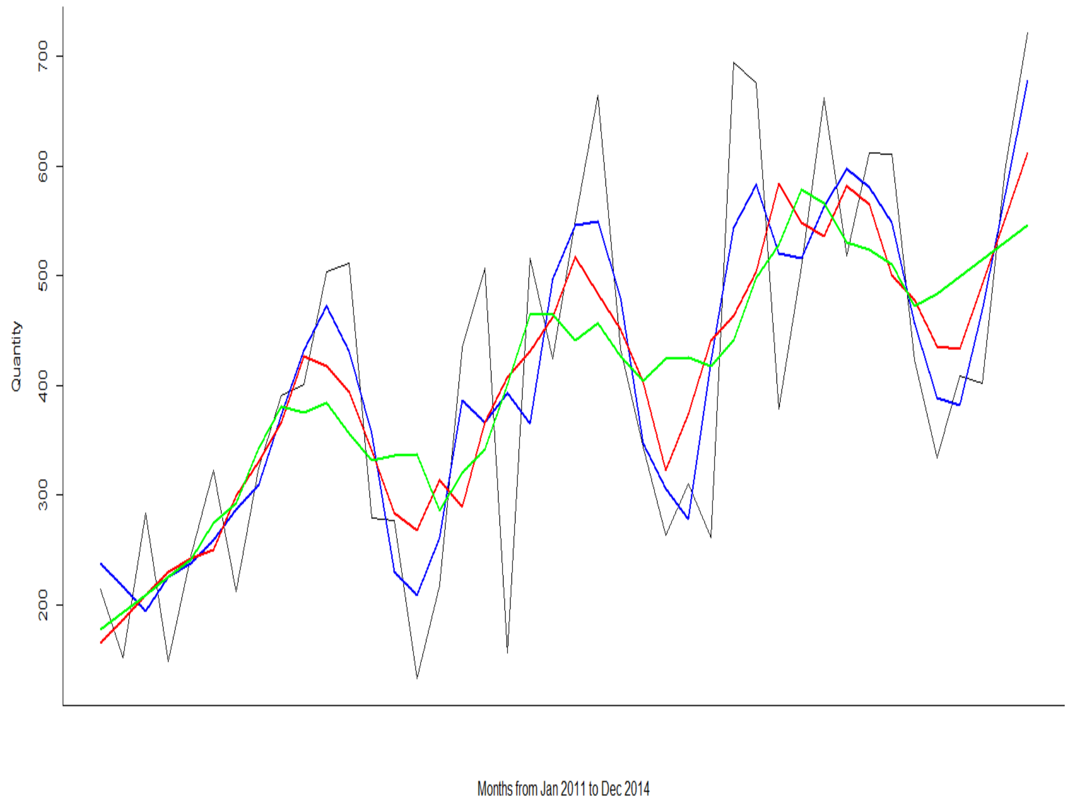


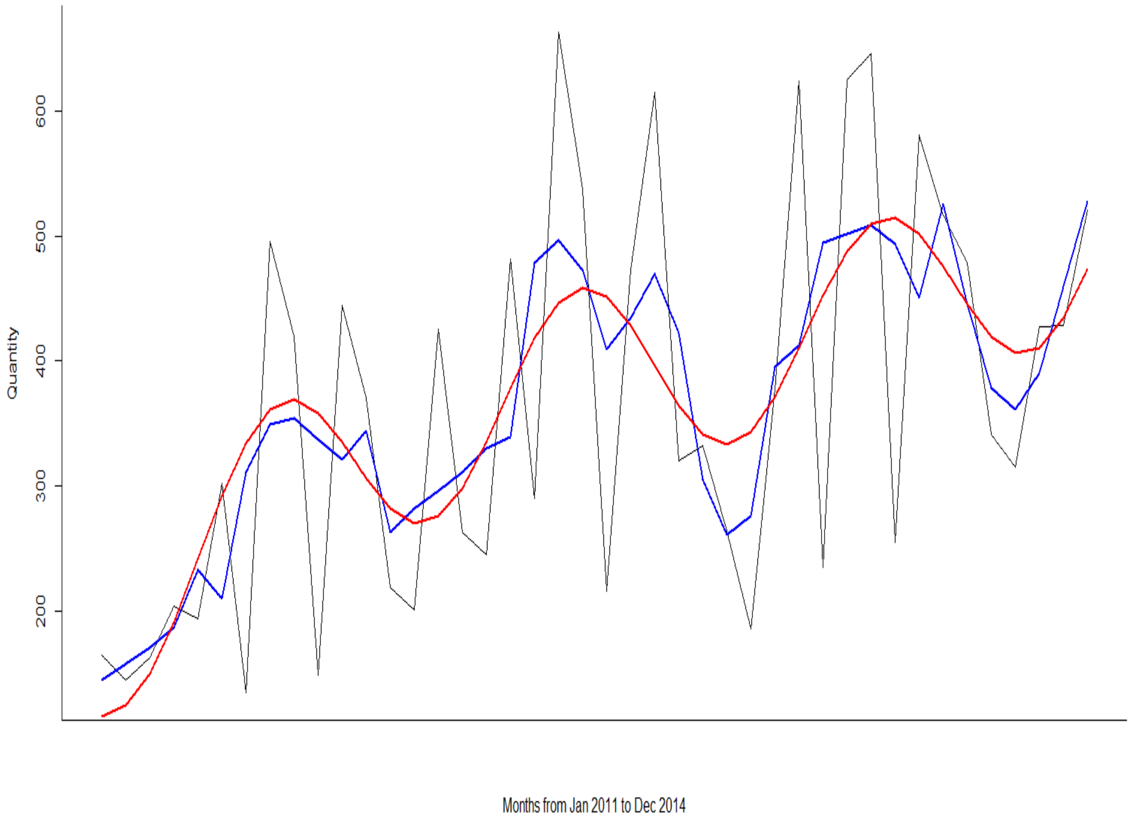
EU Consumer Quantity Time Series



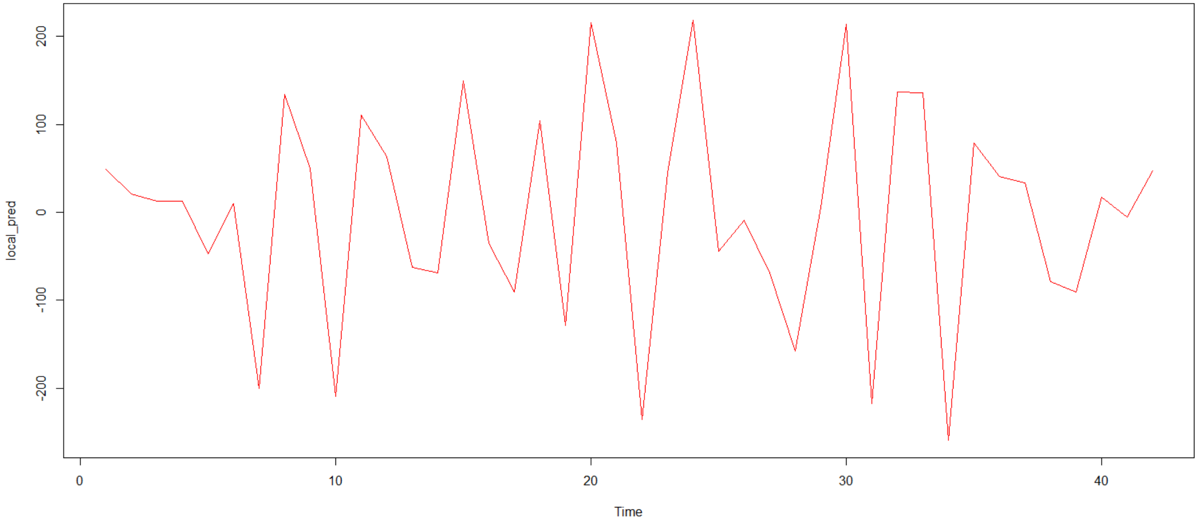


Smoothening for EU Consumer Quantity Time Series

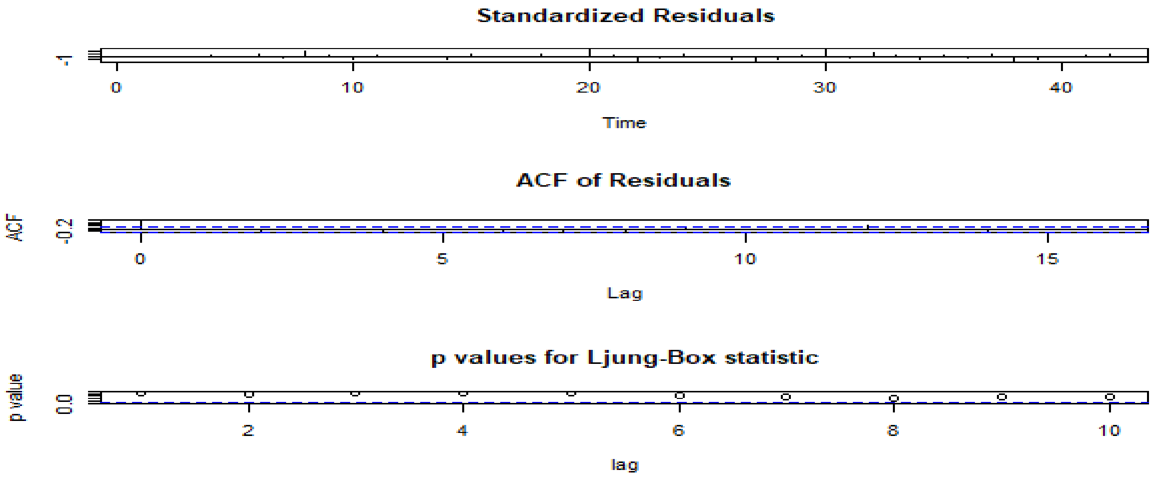


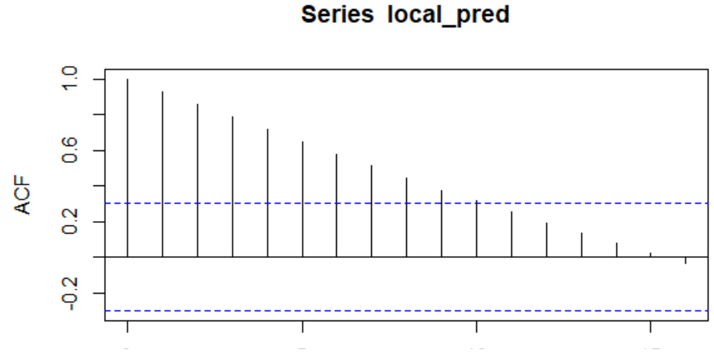


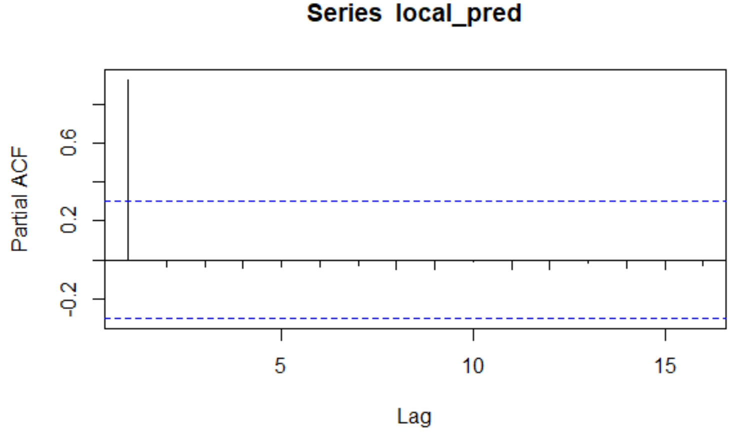
ARMA SERIES for EU Consumer Quantity Time Series



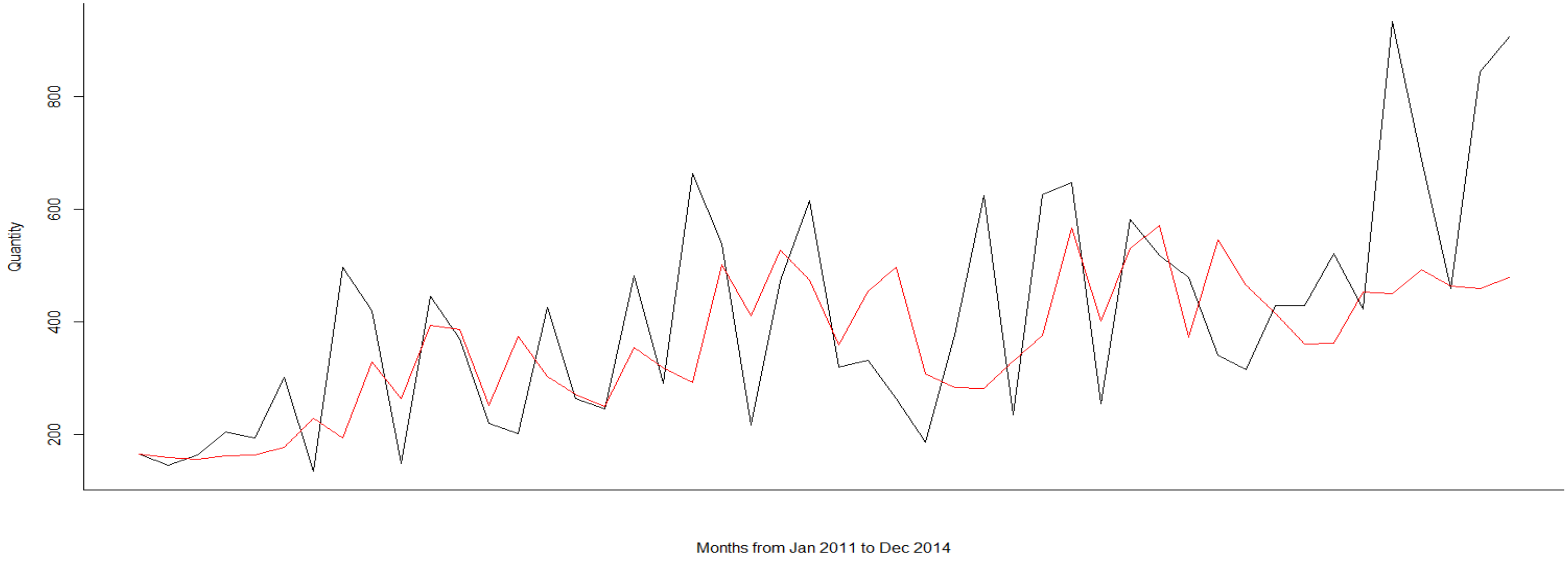
Series: local_pred ARIMA(2,1,0)
log likelihood=-261.57 AIC=531.13
AICc=532.24 BIC=537.99







Forecasting EU Consumer Quantity using ARIMA



Conclusions and Recommendations

- The APAC customer and EU customer segments are the most profitable and consistent segments and hence the management should invest in these two segments.
- For APAC customer segment, multiplicative model gave better results, whereas for EU customer segment, additive model gave better results.
- The MAPE values of all the models are between 24 and 30, which is suitable for forecasting and hence the models are deemed to be appropriate for the analysis.
- Following the expected trend, it is clear from the forecasting plots that the sales are expected to grow.
- But, in the next 6 months, the sales and quantity may get affected because of seasonality.
- Thus, sufficient inventory needs to be accommodated in these two segments.