ECO 5375-701 Prof. Tom Fomby

Eco and Bus Forecasting Fall 2021

**EXERCISE 4**

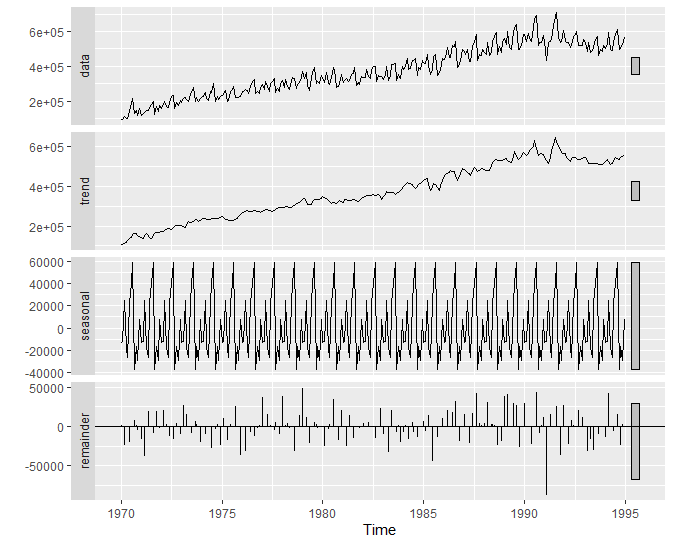
**Purpose:** To learn how to use the stl and stlf function in R. This exercise is due **Tuesday, September 14 at 6:30 pm CST. Submit your work on Canvas.**

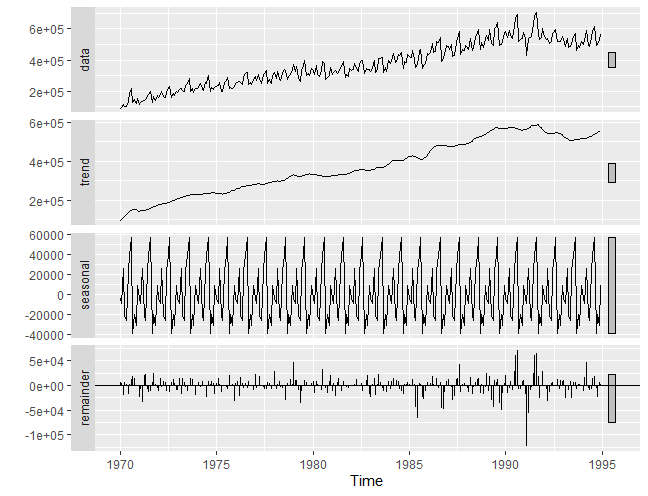
You are to use the Hawaiian data set **Hawaii.csv.** You can use the “starter” R program **EX3\_Hawaiian\_Plots.R** to see how to import the Hawaiian data into RStudio. You are to follow the lead of the R program stl\_stlf demo.R as a guide for completing the following tasks.

Create the training data set, say, hawaii1, and test data set, say, hawaii2. Reserve the first 25 years of data for the training data set (1970:1 – 1994:12) and the last **12** observations for the test data set (1995:1 – 1995:12). Use the “head” and “tail” commands to do this.

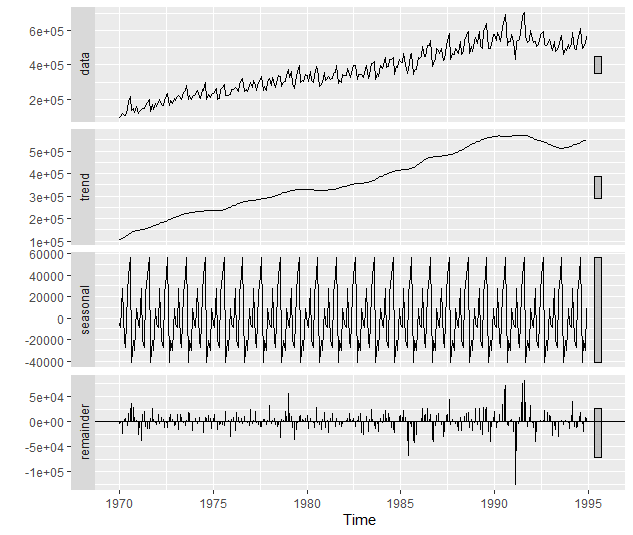
(a) Produce 3 stl time series decompositions using hawaii1, t.window = 5, t.window = 11, and t.window = 21. Report these three stl time series decompositions here.

**t.window = 5**



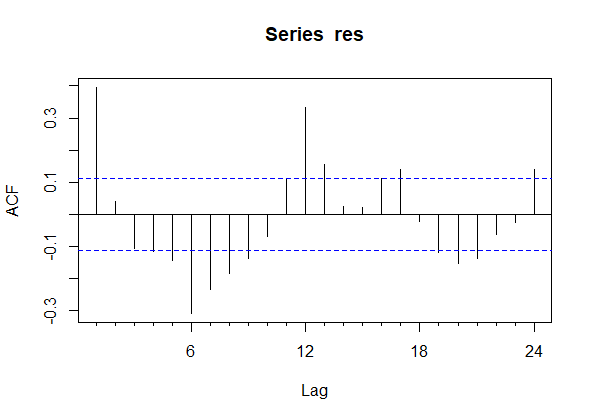
**t.window = 11**

**t.window = 21**

****

(b) Assume the appropriate t.window = 21 as determined by the domain expert. Get the fit of the stl model and report the Acf graph and Box.test results here. What is your conclusion concerning the remainder term of the stl decomposition?

**Acf Graph**



**Box.test results**

Box-Ljung test

data: res

X-squared = 208.76, df = 24, p-value < 2.2e-16

The p-value of the Box-Ljung test is less than the threshold for statical significance (0.05). Therefore, we reject the null hypothesis that the remainder terms are white noise and accept the alternative that they are not white noise.

(c) Get the three test-data set forecasting accuracy results for the “naïve” method used in the stlf function, for the “ets” method, and for the “arima” method. Report the **test data set** results below:

Method RMSE MAE MAPE

naïve 25607.57 21193.83 3.867397

ets 26912.25 23100.37 3.921503

arima 23942.36 19446.8 3.550651

Which is the best performing stlf procedure?

Arima

(d) One more accuracy check. Use the Arima “Airline” model to forecast the test data. What are your **test data set** accuracy results? Report them below.

Method RMSE MAE MAPE

Airline 13307.45 11212.64 2.022189

(e) Of the four methods, which one preformed best? Explain your conclusion.

The Arima “Airline” model performed the best. The Airline model had the lowest RMSE, MAE, and MAPE of all the models we used to forecast the test data.

(e) **Use a block copy to report the R code that you wrote to complete the above tasks.**

library(fpp2)

#setting working directory

setwd("C:/Users/madat/Fall 2021/Assignments/Forecasting/Exercise 3")

# Read in the data

hawaii <- ts(scan("Hawaii.csv"),frequency=12,start=c(1970,1))

# Viewing entire data set

hawaii

# Creating and Viewing the data partitions

# Training data set

hawaii1 <- head(hawaii, 12\*25)

hawaii1

# Test data set

hawaii2 <- tail(hawaii, 12)

hawaii2

# Part A

## STL with window 5

hawaii1 %>%

stl(t.window=5, s.window="periodic", robust=TRUE) %>%

autoplot()

#STL with window 11

hawaii1 %>%

stl(t.window=11, s.window="periodic", robust=TRUE) %>%

autoplot()

#STL with window 21

hawaii1 %>%

stl(t.window=21, s.window="periodic", robust=TRUE) %>%

autoplot()

# Part B

#we are assuming that the appropriate window is 21

fit <- stl(hawaii1, t.window=21, s.window="periodic",

robust=TRUE)

# Displaying the components of the stl decomposition

fit

# Here we are displaying the remainder separately.

remainder(fit)

res <- remainder(fit)

Acf(res)

#Box test

Box.test(res,lag=24,type=c("Ljung-Box"))

# Part C

# In this part of the program we look at three different

# stl forecasts, the three being "naive", "ets", and "arima".

# The forecasts are out-of-sample on the reserved (test) 15

# observations.

forc1 <- stlf(hawaii1, t.window=21, s.window="periodic",

robust=TRUE, method="naive",h=15)

autoplot(forc1)

forc1

# Getting the forecasting accuracy measures for forc1

accuracy(forc1,hawaii2)

forc2 <- stlf(hawaii1, t.window=21, s.window="periodic", robust=TRUE, method="ets",h=15)

autoplot(forc2)

forc2

# Getting the forecasting accuracy measures for forc2

accuracy(forc2,hawaii2)

forc3 <- stlf(hawaii1, t.window=21, s.window="periodic",

robust=TRUE, method="arima",h=15)

autoplot(forc3)

forc3

# Getting the forecasting accuracy measures for forc3

accuracy(forc3,hawaii2)

# Making matrix to display 3 forecasts and test observations.

mat <- c(forc1$mean,forc3$mean,hawaii2)

output <- matrix(mat,nrow=42,ncol=3,byrow=FALSE)

# Putting in the column names for the matrix

colnames(output) <- c("forc1","forc3","hawaii2")

# Part D

forc4 <- forecast(Arima(hawaii1,order=c(0,1,1),

seasonal=c(0,1,1)),h=15)

accuracy(forc4,hawaii2)

summary(forc4)