Homework 3 Electricity Dataset

**#R Code**

**#Load the data & create time series**

library(fpp)

setwd("~/R/PREDICT 413/Homework 3")

EC<- read.csv("Electricity.csv")

ECtimeseries <- ts(EC$Usage, frequency = 12, start= c(1971,1))

ECtimeseries

**#EDA**

summary(EC)

str(EC)

**#Timeplot on Full Data**

par(mfrow=c(2,1))

plot(ECtimeseries, ylab= "Usage", xlab= "Year", main= " Monthly Average Residential Electricity Usage for Iowa City 1971 – 1979")

**#Seasonal Subseries Plot**

monthplot(ECtimeseries,ylab="Usage",xlab="Month",xaxt="n",

main="Seasonal deviation plot: Electricity Usage")

axis(1,at=1:12,labels=month.abb,cex=0.8)

par(mfrow=c(1,1))

**#Stl Decomposition**

fit\_stl <- stl(ECtimeseries, s.window = "periodic")

plot(fit\_stl, main = "STL Decomposition on Full Dataset")

**#Split Data into Train/Test: 80/20**

ECtimeseries\_train <-window(ECtimeseries, frequency = 12, start=c(1971, 1), end=c(1977, 12))

ECtimeseries\_train

ECtimeseries\_test <-window(ECtimeseries, frequency = 12, start=c(1978, 1), end=c(1979, 10))

ECtimeseries\_test

**#Model Selection**

**#ETS Model**

fit1\_ets <- ets(ECtimeseries\_train)

summary(fit1\_ets)

**#Auto.Arima Pre-work**

tsdisplay(ECtimeseries\_train)

**#Unit Root Tests**

adf.test(ECtimeseries\_train, alternative = "stationary")

kpss.test(ECtimeseries\_train)

**#Number of Differencing Required for Seasonal Data**

ns <- nsdiffs(ECtimeseries\_train)

if(ns > 0) {

xstar <- diff(ECtimeseries\_train,lag=frequency(ECtimeseries\_train),differences=ns)

} else {

xstar <- ECtimeseries\_train

}

ns

**#Differencing**

kpss.test(diff(ECtimeseries\_train))

tsdisplay(diff(ECtimeseries\_train))

**#Auto.Arima Model**

fit2\_arima <- auto.arima(ECtimeseries\_train, stepwise=FALSE, approximation=FALSE)

summary(fit2\_arima)

tsdisplay(residuals(fit2\_arima))

Box.test(residuals(fit2\_arima), fitdf=3, lag=10, type="Ljung")

tsdiag(fit2\_arima)

**#Neutral Net**

fit3\_nn <- nnetar(ECtimeseries\_train)

summary(fit3\_nn)

**#Holt Winters Multiplicative with damp**

fit4\_hw <- hw(ECtimeseries\_train, seasonal = "multiplicative", damped = TRUE)

summary(fit4\_hw)

**#Benchmark**

fit5\_meanf<-meanf(ECtimeseries\_train)

fit6\_naive<-naive(ECtimeseries\_train)

fit7\_rwfdrift<-rwf(ECtimeseries\_train, drift=TRUE)

fit8\_snaive<-snaive(ECtimeseries\_train)

**#Training Set Accuracy - Summary**

summary(fit1\_ets) # training set

summary(fit2\_arima) # training set

summary(fit3\_nn) # training set

summary(fit4\_hw) # training set

summary(fit5\_meanf) # training set

summary(fit6\_naive) # training set

summary(fit7\_rwfdrift) # training set

summary(fit8\_snaive) # training set

**#Training Set Accuracy – Goodness-of-fit**

accuracy(fit1\_ets) # training set

accuracy (fit2\_arima) # training set

accuracy (fit3\_nn) # training set

accuracy (fit4\_hw) # training set

accuracy (fit5\_meanf) # training set

accuracy (fit6\_naive) # training set

accuracy (fit7\_rwfdrift) # training set

accuracy (fit8\_snaive) # training set

**#Forecast on Test Set**

par(mfrow=c(3,2))

ETS\_MNM <-forecast(fit1\_ets, h=length(ECtimeseries\_test))

plot(ETS\_MNM, ylab="Usage")

lines(ECtimeseries, col="red",ylab="Actual")

ETS\_MNM

Auto.ARIMA <-forecast(fit2\_arima, h=length(ECtimeseries\_test))

plot(Auto.ARIMA, ylab="Usage")

lines(ECtimeseries, col="red",ylab="Actual")

Auto.ARIMA

NN <-forecast(fit3\_nn, PI=TRUE, h=length(ECtimeseries\_test))

plot(NN, ylab="Usage")

lines(ECtimeseries, col="red",ylab="Actual")

NN

HW <-forecast(fit4\_hw, h=length(ECtimeseries\_test))

plot(HW, ylab="Usage")

lines(ECtimeseries, col="red",ylab="Actual")

HW

SNAIVE <-snaive(ECtimeseries\_train, h=length(ECtimeseries\_test))

plot(SNAIVE)

lines(ECtimeseries, col="red",ylab="Actual", ylab="Usage")

par(mfrow=c(1,1))

print(accuracy(ETS\_MNM, ECtimeseries\_test)) #best model

print(accuracy(Auto.ARIMA, ECtimeseries\_test))

print(accuracy(NN, ECtimeseries\_test))

print(accuracy(HW, ECtimeseries\_test))

print(accuracy(SNAIVE, ECtimeseries\_test))

**#Diagnostics**

**#Box-Ljung test A**

Box.test(ETS\_MNM$residuals, lag=25, type = "Ljung-Box")

par(mfrow=c(2,2))

acf(ETS\_MNM$residuals, lag.max=25)

plot(ETS\_MNM$residuals, ylab = "Residuals")

abline(h = 0, col = "red")

hist(ETS\_MNM$residuals, main = "", xlab = "Residuals")

par(mfrow=c(1,1))

**#Box-Ljung test B**

Box.test(Auto.ARIMA$residuals, lag=25, type = "Ljung-Box")

par(mfrow=c(2,2))

acf(Auto.ARIMA $residuals, lag.max=25)

plot(Auto.ARIMA $residuals, ylab = "Residuals")

abline(h = 0, col = "red")

hist(Auto.ARIMA $residuals, main = "", xlab = "Residuals")

par(mfrow=c(1,1))

**#Box-Ljung test C**

par(mfrow=c(2,2))

plot(NN$residuals, ylab = "Residuals")

abline(h = 0, col = "red")

hist(NN$residuals, main = "", xlab = "Residuals")

par(mfrow=c(1,1))

**#Box-Ljung test D**

par(mfrow=c(2,2))

plot(HW$residuals, ylab = "Residuals")

abline(h = 0, col = "red")

hist(HW$residuals, main = "", xlab = "Residuals")

par(mfrow=c(1,1))

**#Forecast of Next 5 Months**

par(mfrow=c(3,2))

fit3 <- ets (ECtimeseries, model ="MNM")

ETS\_MNM <-forecast(fit3, h=5)

ETS\_MNM

plot(ETS\_MNM, ylab="Usage")

fit4 <- Arima(ECtimeseries, order=c(1,0,0), seasonal=c(0,1,1), include.drift=TRUE)

Auto.ARIMA <-forecast(fit4, h=5)

plot(Auto.ARIMA, ylab="Usage")

Auto.ARIMA

fit5 <- nnetar(ECtimeseries, order=c(1,1,2))

NN <-forecast(fit5, PI=TRUE, h=5)

plot(NN, ylab="Usage")

NN

fit6 <- hw(ECtimeseries, seasonal = "multiplicative", damped = TRUE)

HW <-forecast(fit6, h=5)

plot(HW, ylab="Usage")

HW

fit7 <- snaive (ECtimeseries, h=5)

plot(fit7)

par(mfrow=c(1,1))