#### Consumer ####

#### Problem 1 ####

# p 94, green Tsay

# my Problem 1

library(tseries)

library(fBasics)

library(forecast)

library(lmtest)

da=read.table("~/Desktop/TIme Series/TS3/m-umcsent.txt",header=T)

head(da)

csent=da$VALUE

# here I made the csent into a formal time series so I could use other time series functions

csentts <-ts(csent,start = 1978,freq=12)

plot(csentts)

qqnormPlot(csentts)

hist(csentts,prob =T, ylim= c(0,.05),col= "red")

lines(density(csentts),lwd=2)

shapiro.test(csentts)

# The shapiro test reveal that either the csentts or lgcsentts is a normal distribution

# Due to the variance the data is log

lgcsentts <- log(csentts +1)

plot(lgcsentts)

hist(lgcsentts,prob =T, ylim= c(0,5),col= "orange")

lines(density(lgcsentts),lwd=2)

shapiro.test(lgcsentts)

#(a) Conduct a complete EDA on the consumer data

# Look at basic stats

basicStats(lgcsentts)

qqnormPlot(lgcsentts)

# The basic stats and qqnormplot reveals a non stationary process with skewness and a non str line

# Next Graph the time series

tsdisplay(lgcsentts)

t.test(lgcsentts)

# With the Graph displaying up trends and down trend it is not stationary but hard to tell if there is a seasonal aspect

#The t.test reveals the mean is signficantly different than 0

# Also the ACF plot reveals reveals high sample serial correlation

# I will look at differencing the time series

tsdisplay(diff(lgcsentts)) # Note PACF at 10 is significant. Will use that in following test.

# (b) Are there unit roots?

#Will Do Dickey\_Fuller Test for presence of unit-root. The null hypothesis is that there is a trend

adf.test(lgcsentts,k=4,alternative = "stationary")### select lag = 10.

# so the Augmented Dickey\_Fuller Test reveals a p value of 0.4503 and we can not reject the null

# so we accept the NULL that there is a unit-root presence

# if you change the lag to 1 and 4 you can also reject the null and determine their is a unit root

# so let's do another unit-root test the KPSS test

# For this test the null is there is no unit-root which is opposite the Dickey FUller test

kpss.test(lgcsentts)

# The p value is very low so the null is rejection and thus there is a unit root present in this data.

# Do we need more differening than first-differencing.

ndiffs(lgcsentts)

ndiffs(diff(lgcsentts))

# these two test reveal that we only need one differencing

# do we need any seasonal differencing? The following test will answer that question.

nsdiffs(lgcsentts)

nsdiffs(diff(lgcsentts))

# We do not have any season differencing.

# Let's do stl on the lgcsentts

dec <-stl(lgcsentts, s.window = "periodic")

summary(dec)

plot(dec)

# The stl reveals a large trend and a minor seasonal component

#(c) Do first differencing.

rt<-diff(lgcsentts)

hist(rt,prob =T, ylim= c(0,10),col= "blue")

lines(density(rt),lwd=2)

basicStats(rt)

# mean at almost o and skewness and kurtosis is small

qqnormPlot(rt)

# qq plot reveals a much better straignt line but some outliers are present

tsdisplay(rt)

t.test(rt)

# mean is not different that 0

decom2<- stl(rt, s.window = "periodic")

summary(decom2)

plot(decom2)

# Stl reveals most of the trend has been removed

# not part of the problem but---

# Just to check and see with arima model is produced. Only one trend differencing and no seasonal differencing.

auto.arima(lgcsentts)

auto.arima(rt)