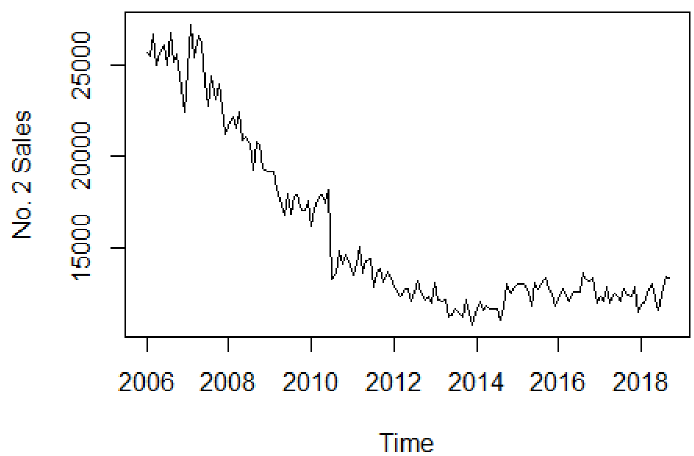
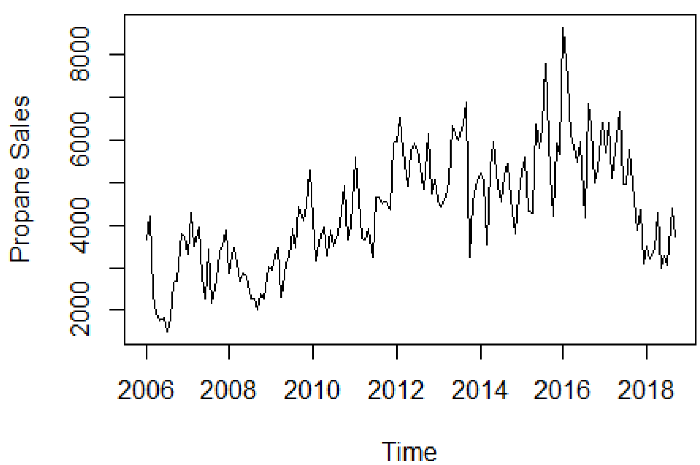
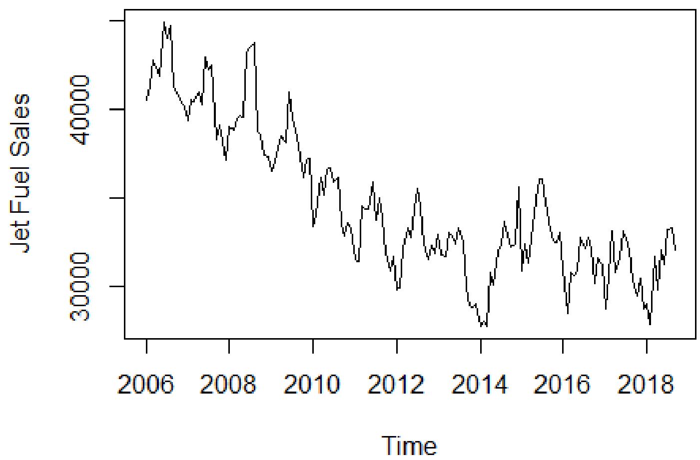
R Code included in Appendix.

Question 1

To begin, I read in the data, initialized the three time series models for each product, and plotted the three.



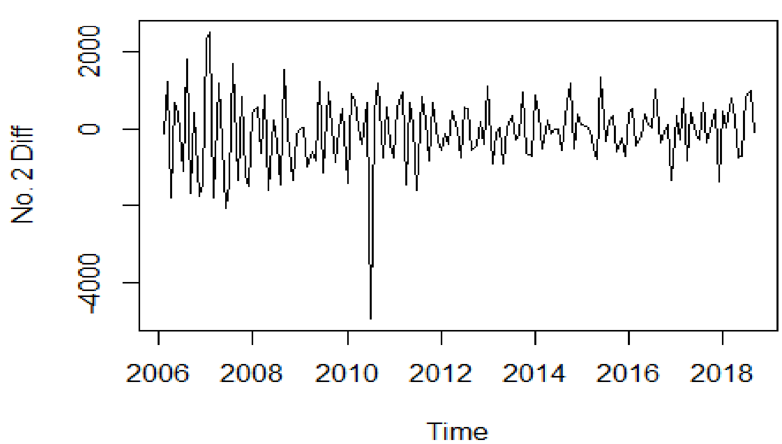


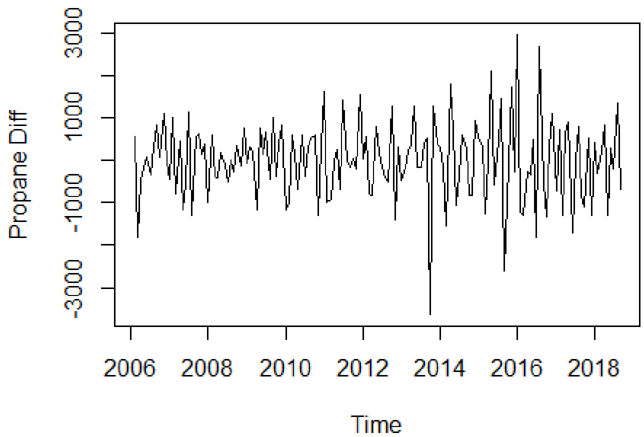


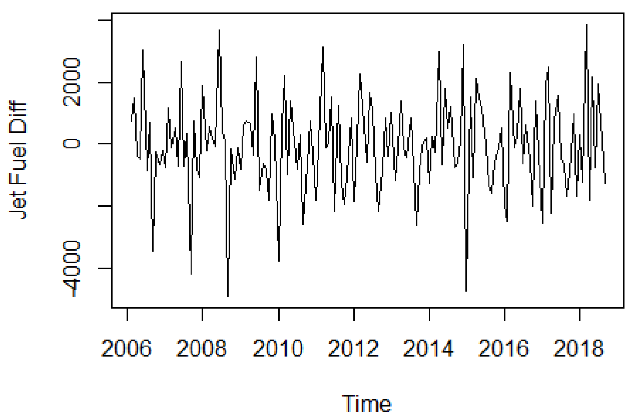
After plotting the three time-series, we can begin to analyze them. The first and most obvious feature in all three series is the presence of a trend. For No. 2 and Jet Fuel, there is a gradual downward trend that levels off as time increases. For Propane, the trend is increasing with time and then seems to decrease. Based on these plots, there could also be an element of a cyclical trend. The sales for all three products seem to rise and fall, but not for a fixed time period. It is also possible that the series exhibit heteroskedasticity, but it is not easily concluded from the plots.

Using the plots, we can also see if the assumptions of stationarity are violated. In all three cases, the assumption of constant mean is clearly violated. There is also a suggestion of non-constant variance/covariance, but again it is not clearly defined. Propane seems to have the most dynamic variance over time.

Next, I analyzed the 1st order differenced time series plot of the data.





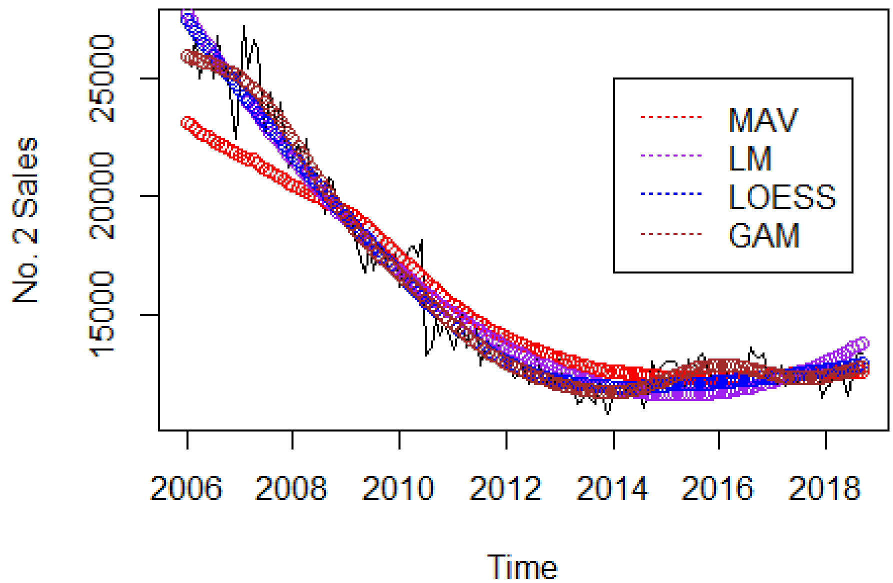


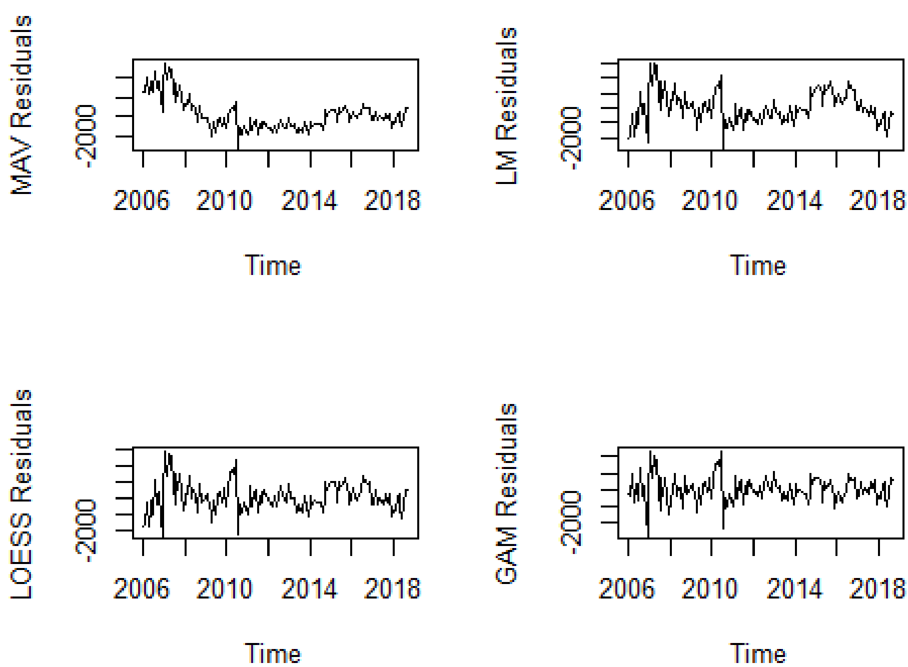
The first order differenced time series show very different plots than the originals. For all three products, any evidence of a general trend is now gone. There could still be a cyclical trend, though. It is also much easier to see heteroskedasticity now. All three series seem to show changing variance at certain times.

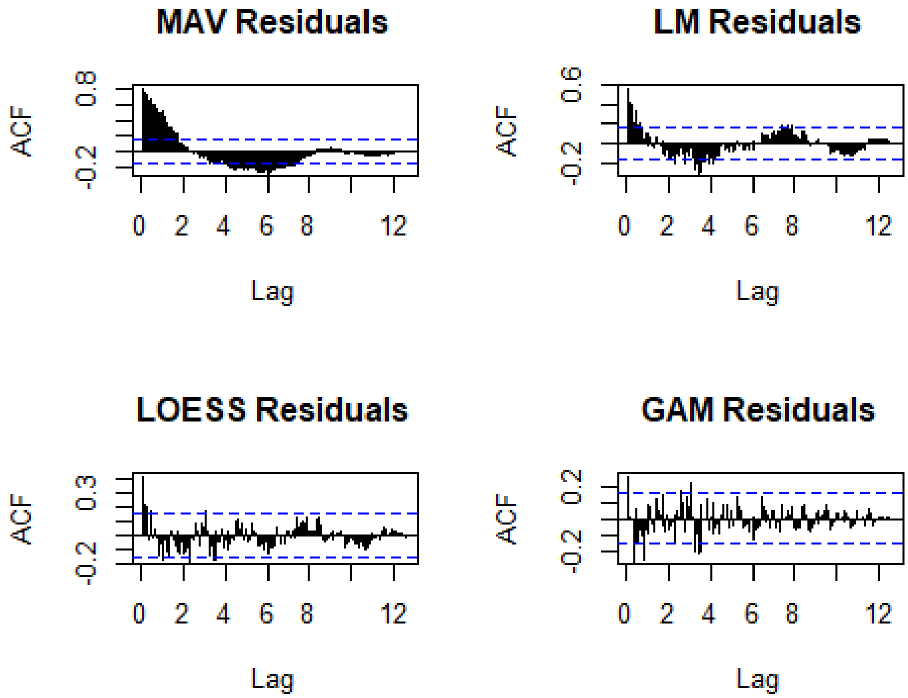
Again, we check if the assumptions of stationarity are violated. There now appears to constant means, roughly centered around zero. So that assumption is no longer violated. The variance for all three appears to be finite, although No.2 has one very large downward spike that may represent an outlier and a large variance. The assumption of constant variance/covariance does appear to be violated by these plots.

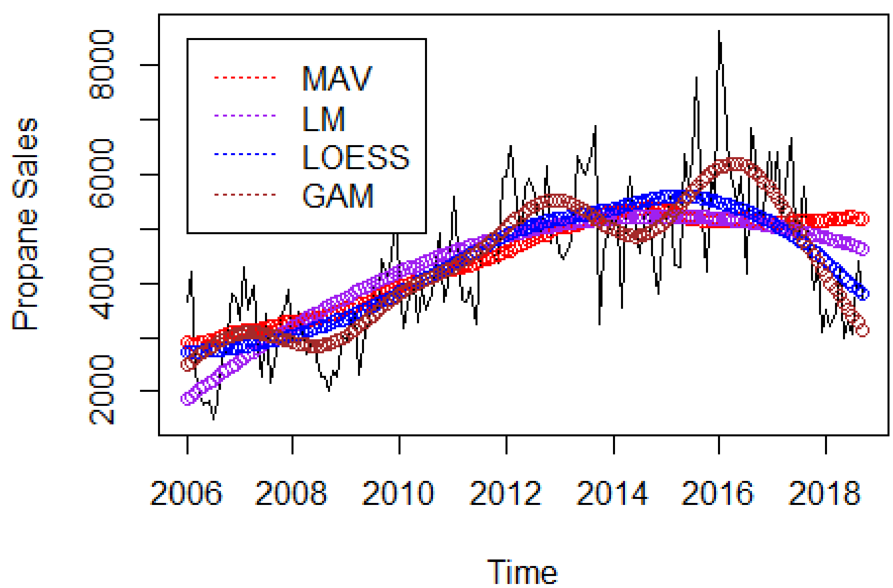
Question 2

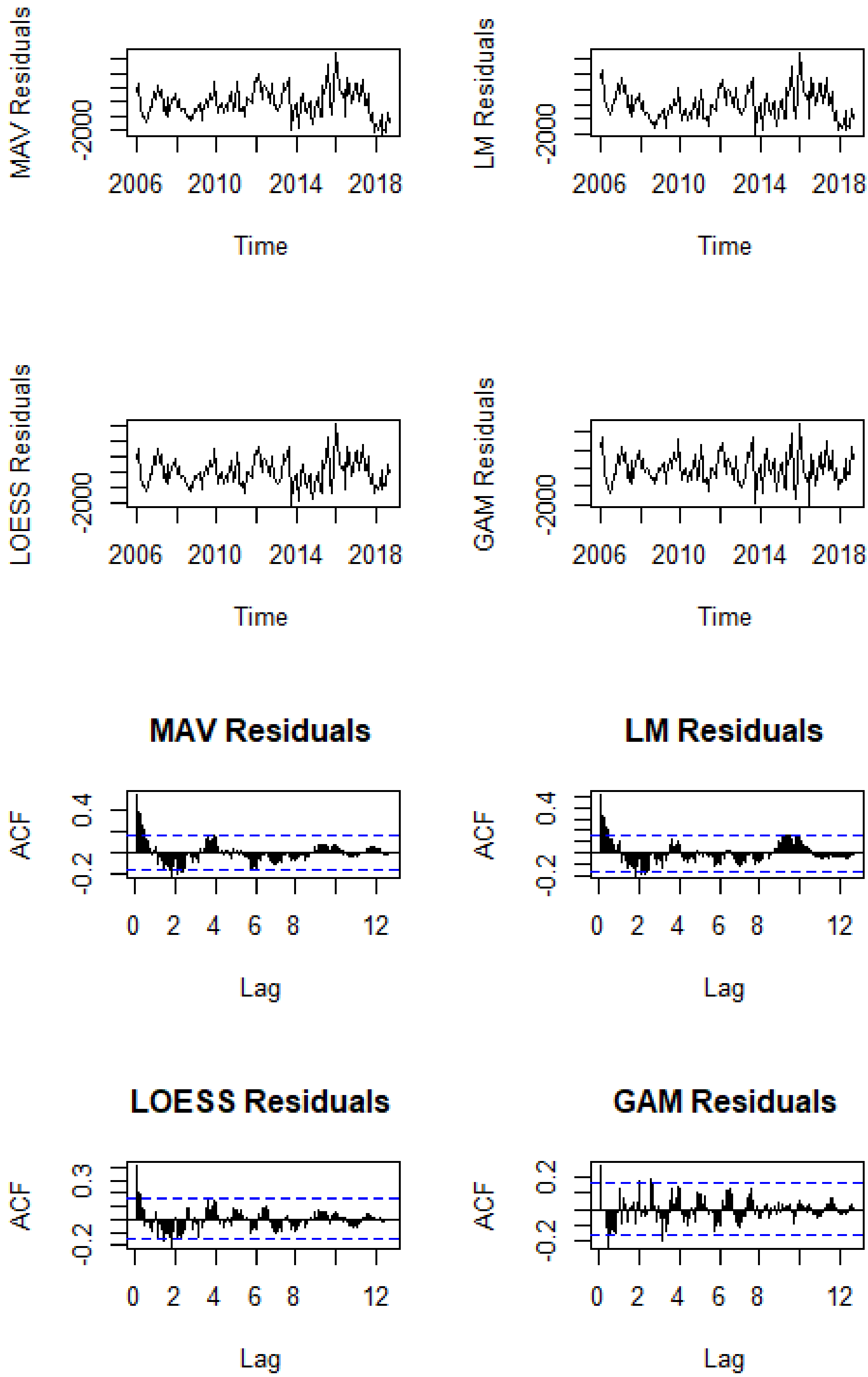
For this question, I attempted to estimate the trend in sales for the three products using different models. For each product, I fit a moving average, parametric quadratic polynomial, local polynomial, and splines model. I then plot the fitted values vs. originals along with the residuals and ACF of the residuals for each model and product.

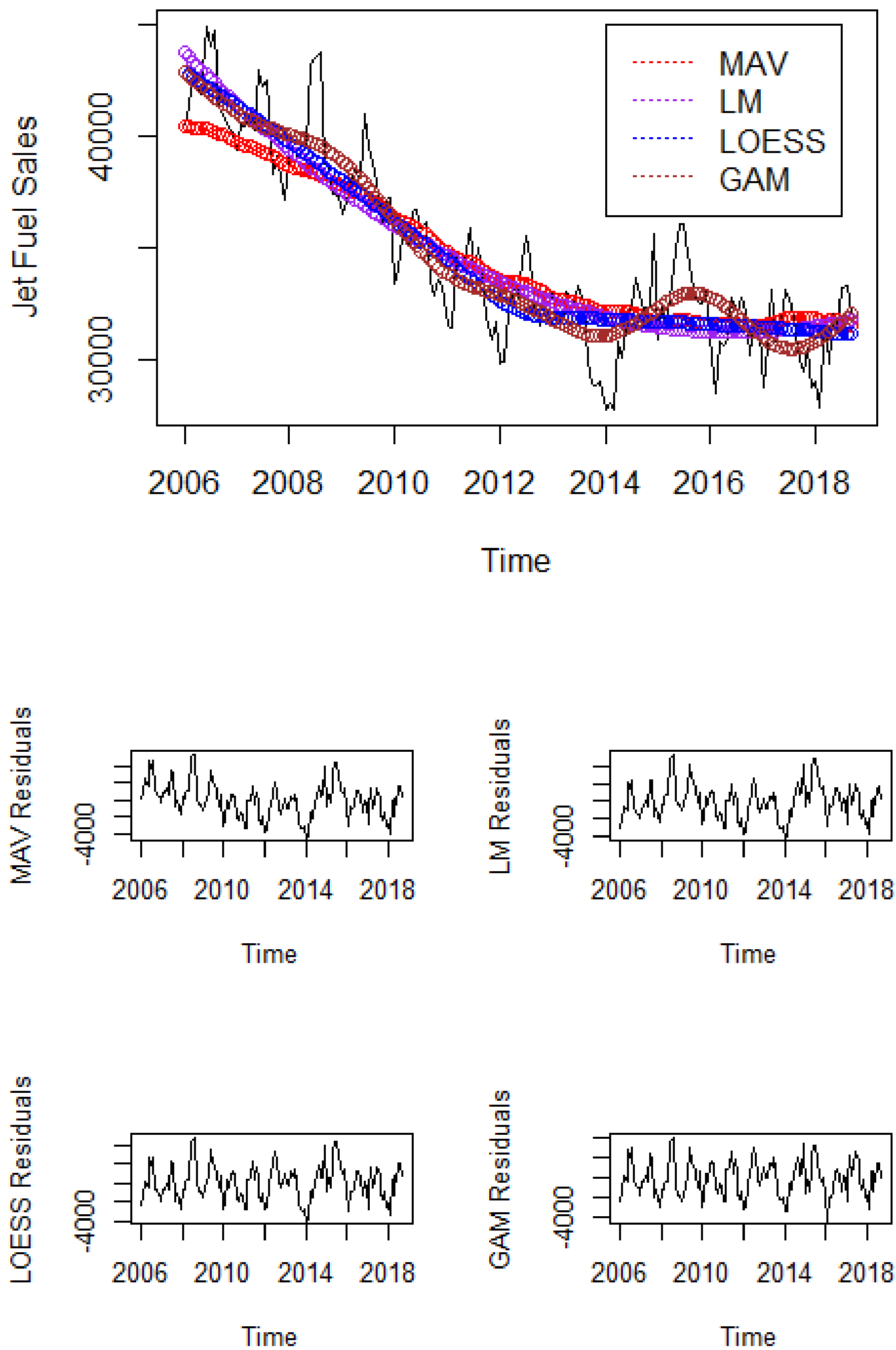


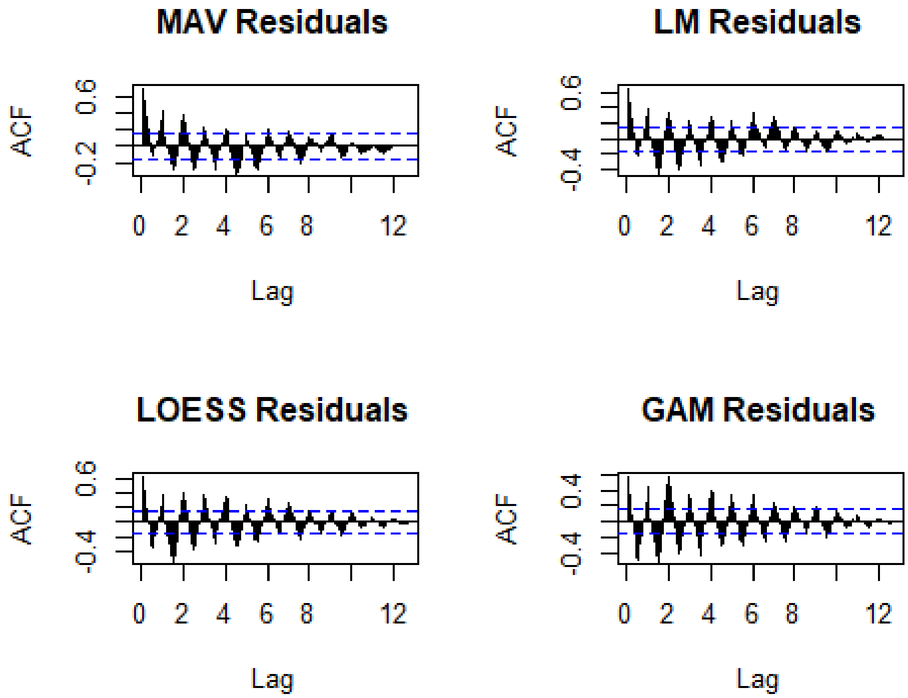



MAV Residuals
LM Residuals
LOESS Residuals
GAM Residuals
ACF
Lag


MAV
LM
LOESS
GAM
Propane Sales
Time


MAV Residuals
-2000
2006
2010
2014
2018
Time
LM Residuals
-2000
2006
2010
2014
2018
Time
LOESS Residuals
-2000
2006
2010
2014
2018
Time
GAM Residuals
-2000
2006
2010
2014
2018
Time
MAV Residuals
ACF
0.4
-0.2
0
2
4
6
8
12
Lag
LM Residuals
ACF
0.4
-0.2
0
2
4
6
8
12
Lag
LOESS Residuals
ACF
0.3
-0.2
0
2
4
6
8
12
Lag
GAM Residuals
ACF
0.2
-0.2
0
2
4
6
8
12
Lag


MAV
LM
LOESS
GAM
Jet Fuel Sales
40000
30000
2006
2008
2010
2012
2014
2016
2018
Time
MAV Residuals
LM Residuals
LOESS Residuals
GAM Residuals
-4000

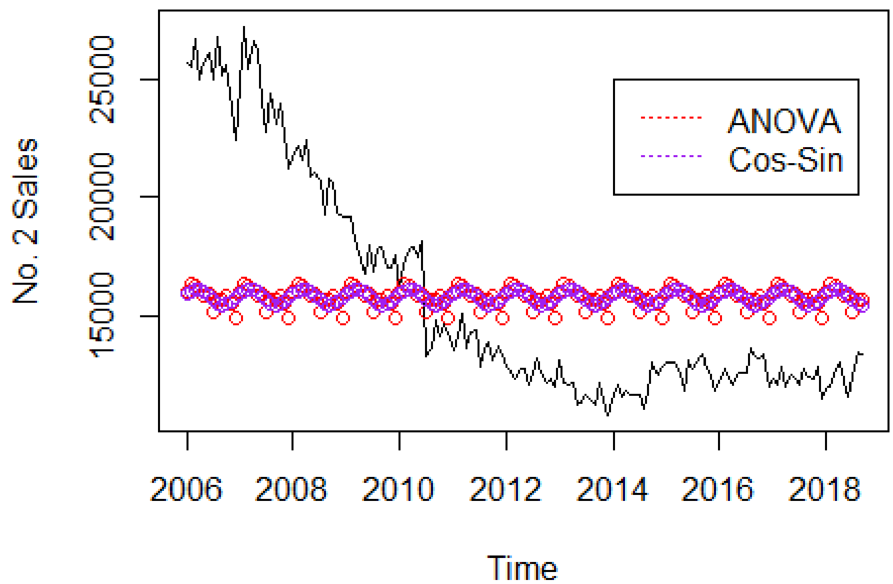


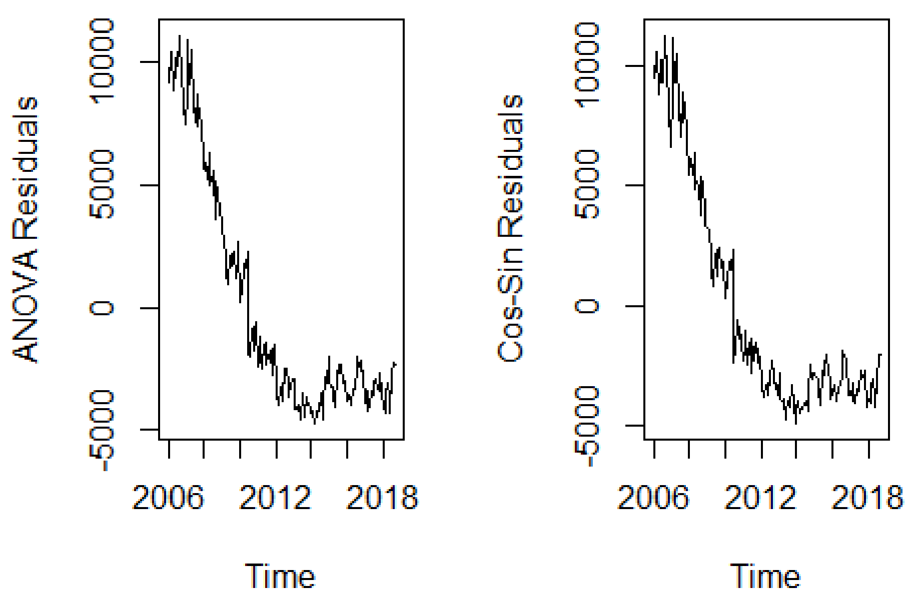
In the original time series plots, there was a clear trend in the data. Therefore, it is not surprising that for each product, the four different trend models fit the data well. As can be seen on the fitted value plots, the models all capture the general direction of the trend for each product. Although not overly obvious, the MAV models seemed to have the worst fit, while the non-parametric models fit the best.

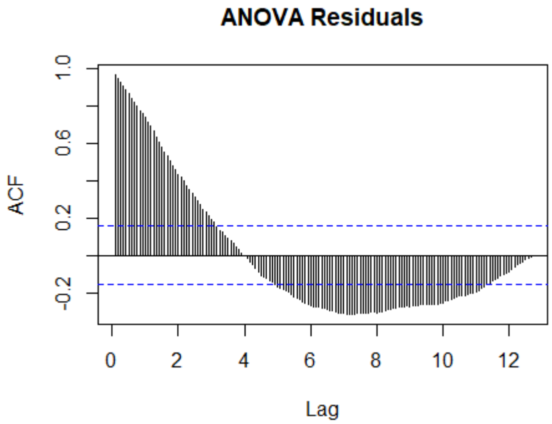
An analysis of the residual plots and ACF of the residuals help us evaluate stationarity again. For all three products (No. 2, Propane, and Jet Fuel) all four models' residuals are roughly centered around zero. However, there is still evidence of a cyclical trend. For Jet Fuel, these trends are starting to look like seasonality. To dig deeper, we look at the ACF plots. Indeed, all three products show evidence of a cyclical trend and Jet Fuel shows strong evidence of a seasonal trend. Therefore, the residuals for all of the models cannot be considered stationary.

Question 3

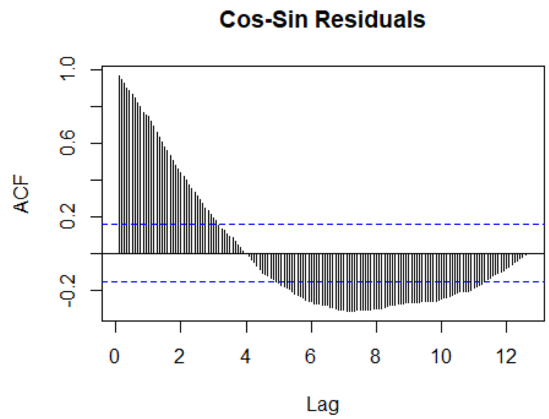
For this question, I attempted to estimate the seasonality in sales for the three products using different models. For each product, I fit a seasonal means (ANOVA) and Cos-Sin model. I then plot the fitted values vs. originals along with the residuals and ACF of the residuals for each model and product.


ANOVA
Cos-Sin
No. 2 Sales
Time
2006
2008
2010
2012
2014
2016
2018
15000
20000
25000

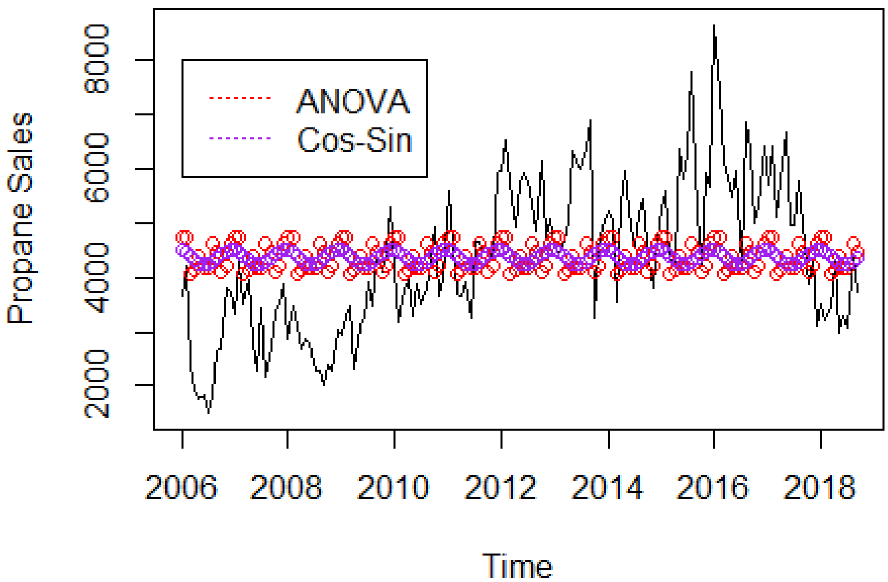

ANOVA Residuals
Cos-Sin Residuals
Time
Time
2006
2012
2018
-5000
0
5000
10000



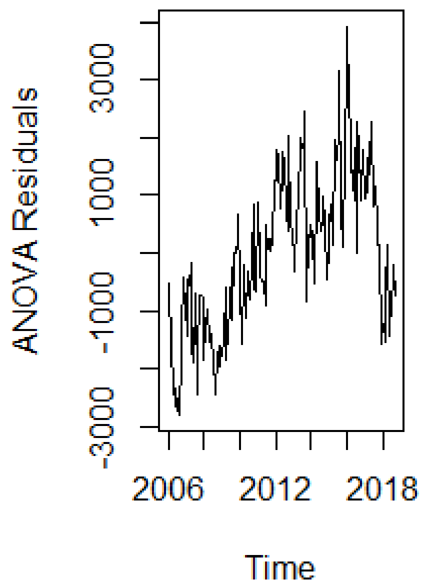
ANOVA Residuals
ACF
1.0
0.6
0.2
-0.2
0
2
4
6
8
10
12
Lag

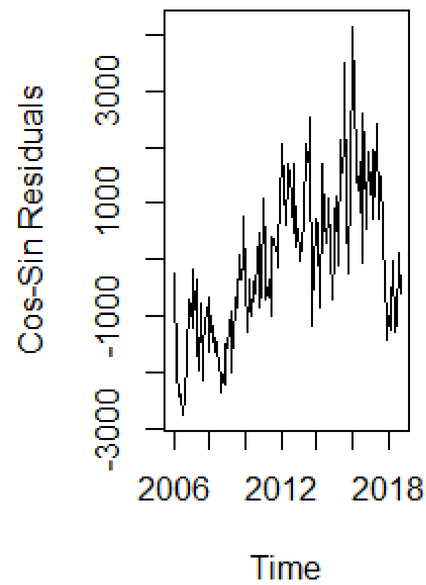


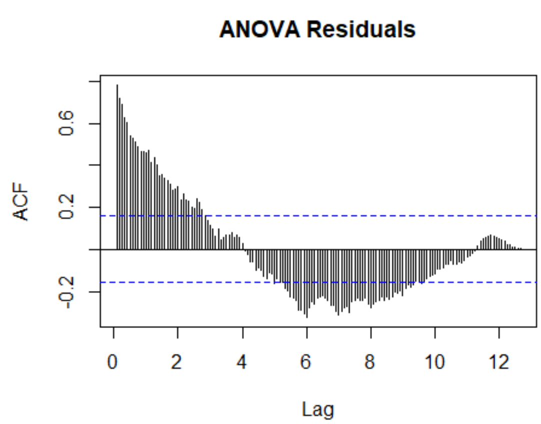
Cos-Sin Residuals
ACF
1.0
0.6
0.2
-0.2
0
2
4
6
8
10
12
Lag

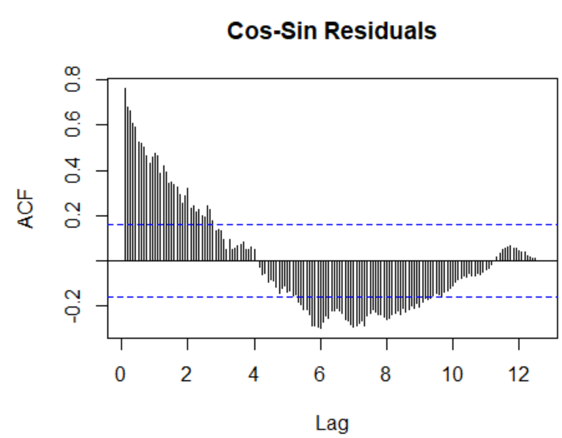


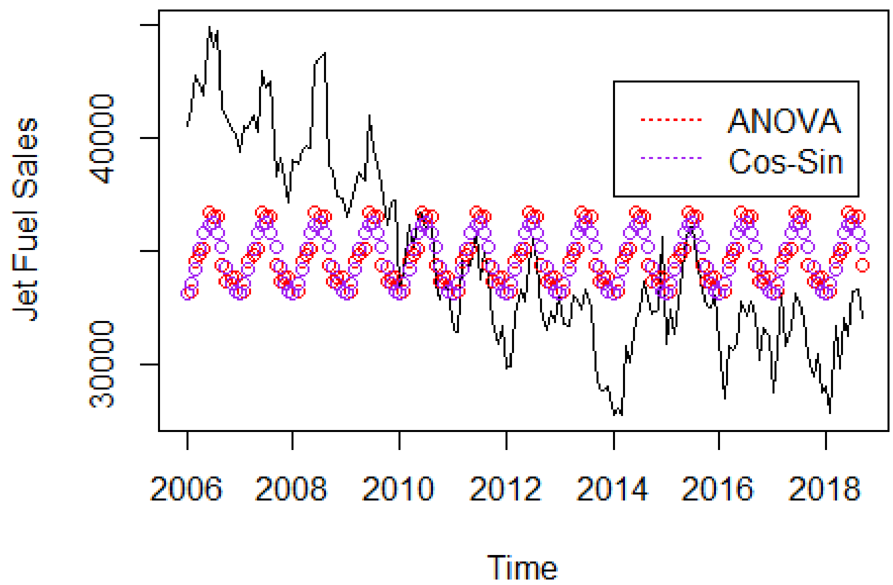
ANOVA
Cos-Sin
Propane Sales
8000
6000
4000
2000
2006
2008
2010
2012
2014
2016
2018
Time

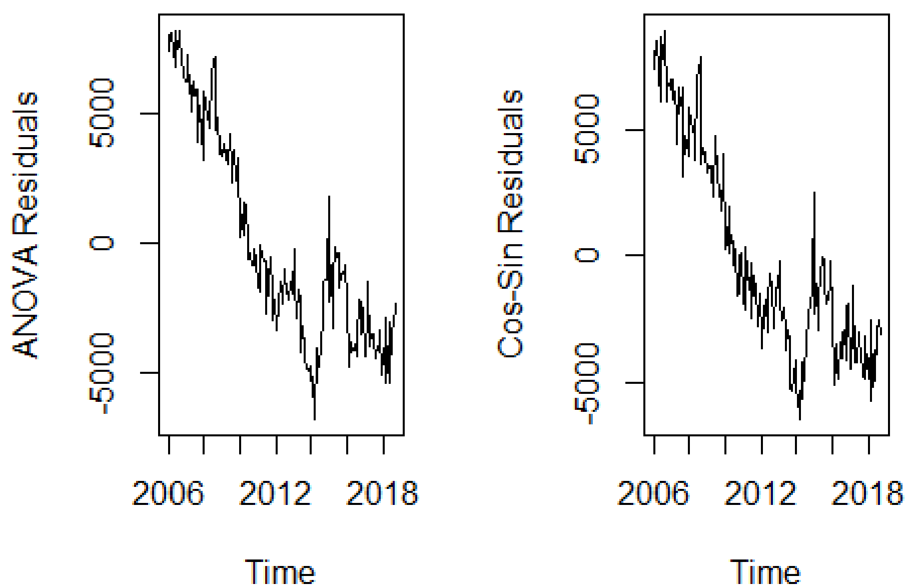

ANOVA Residuals
-3000
-1000
1000
3000
2006
2012
2018
Time

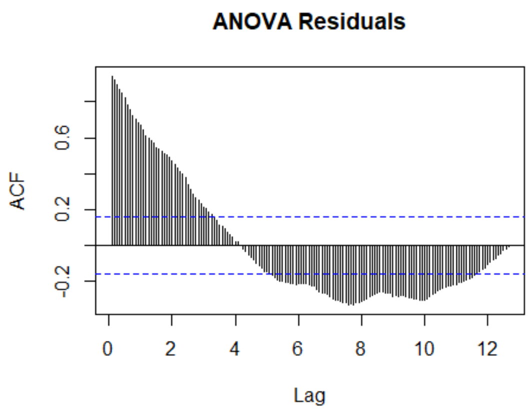

Cos-Sin Residuals
-3000
-1000
1000
3000
2006
2012
2018
Time

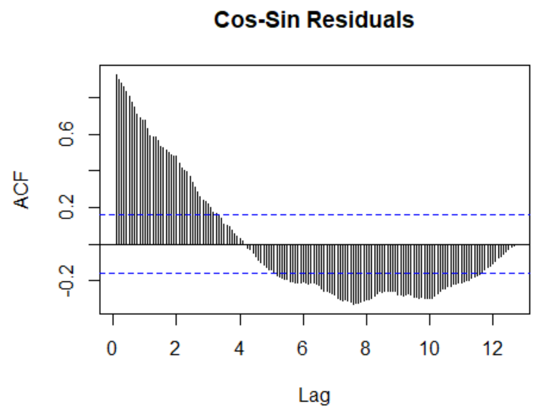

ANOVA Residuals
ACF
-0.2
0.2
0.6
0
2
4
6
8
10
12
Lag


Cos-Sin Residuals
ACF
-0.2
0.2
0.4
0.6
0.8
0
2
4
6
8
10
12
Lag


Jet Fuel Sales
30000
40000
2006
2008
2010
2012
2014
2016
2018
Time
ANOVA
Cos-Sin


ANOVA Residuals
-5000
0
5000
2006
2012
2018
Time
Cos-Sin Residuals
-5000
0
5000
2006
2012
2018
Time



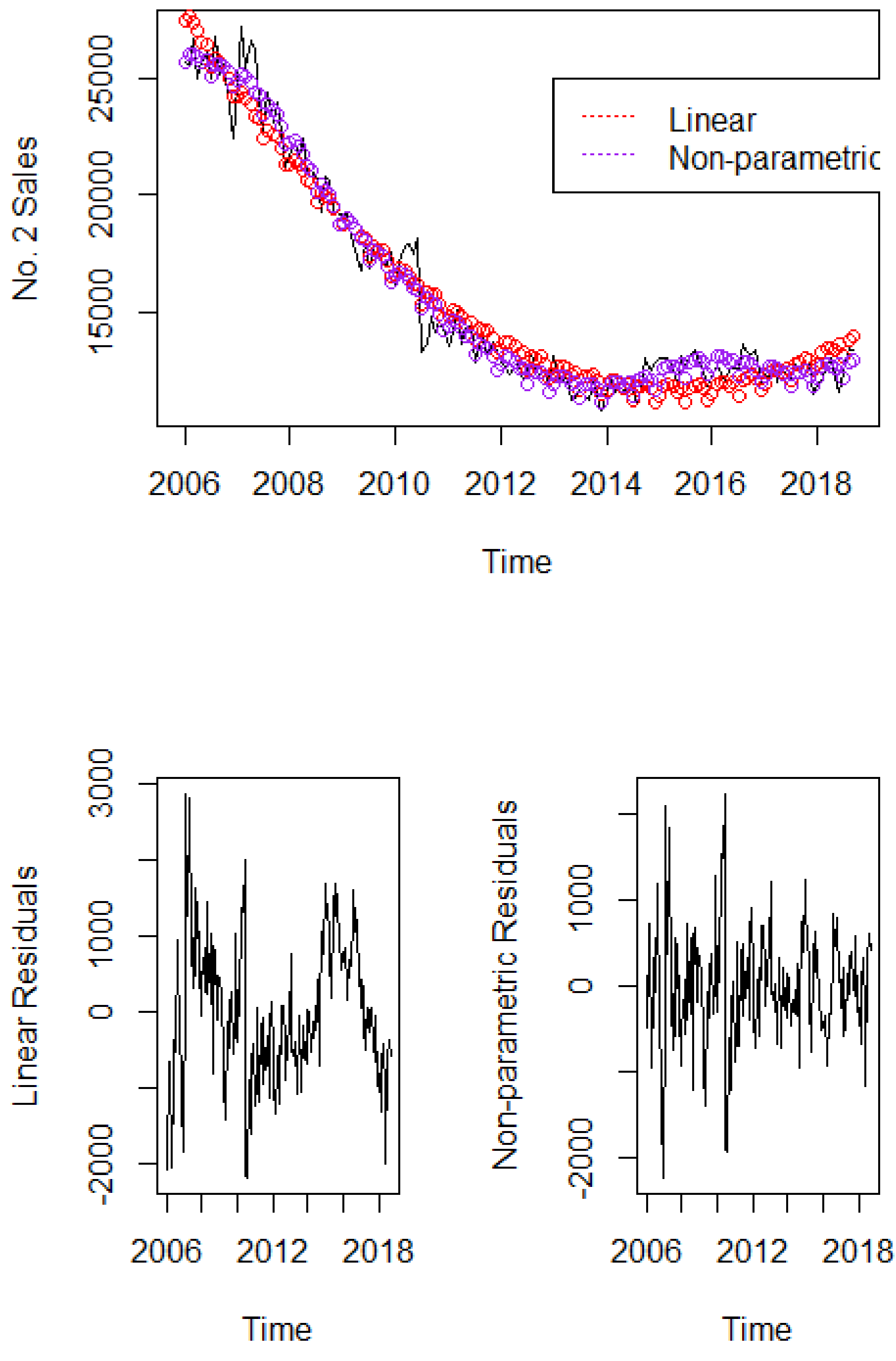


After fitting all of my seasonality models, I again looked at both model fit and stationarity. The fit of these models is harder to visualize with the plots, but looking closely, the models appear to loosely follow rises and falls in the data. Both the ANOVA (seasonal means) and Cos-Sin models fit the data in similar ways.

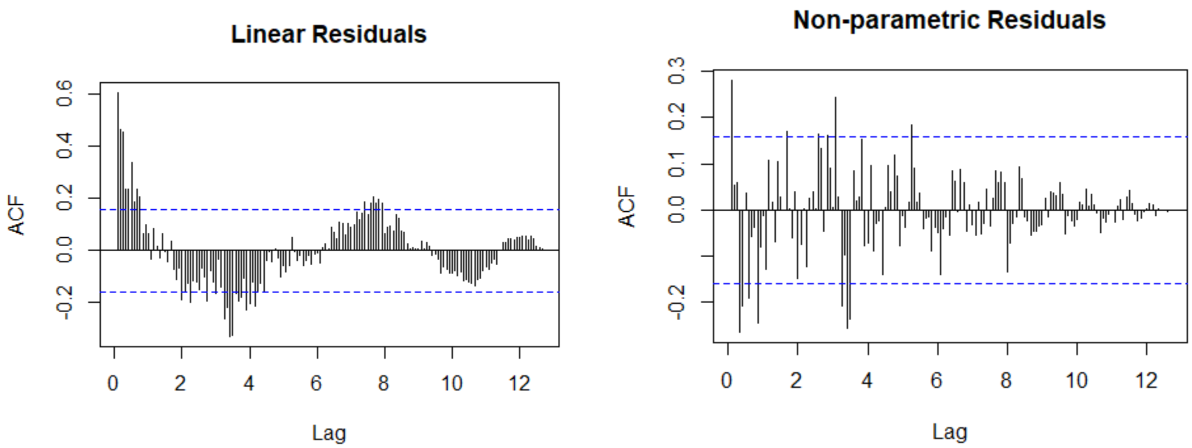
As before, I looked at the residual plots and ACF plots to evaluate stationarity. The residual plots still show the trends that we know exist from the previous question since we did not remove them in this case. We still see the decreasing trend for No. 2 and Jet Fuel, and the increasing trend for Propane. The ACF plots, however, reveal that the evidence of cyclical trends or seasonality, has been greatly reduced. Overall, these residuals cannot be considered stationary due to the presence of the trend. The residuals are not mean constant and seem to have changing variance.

Question 4

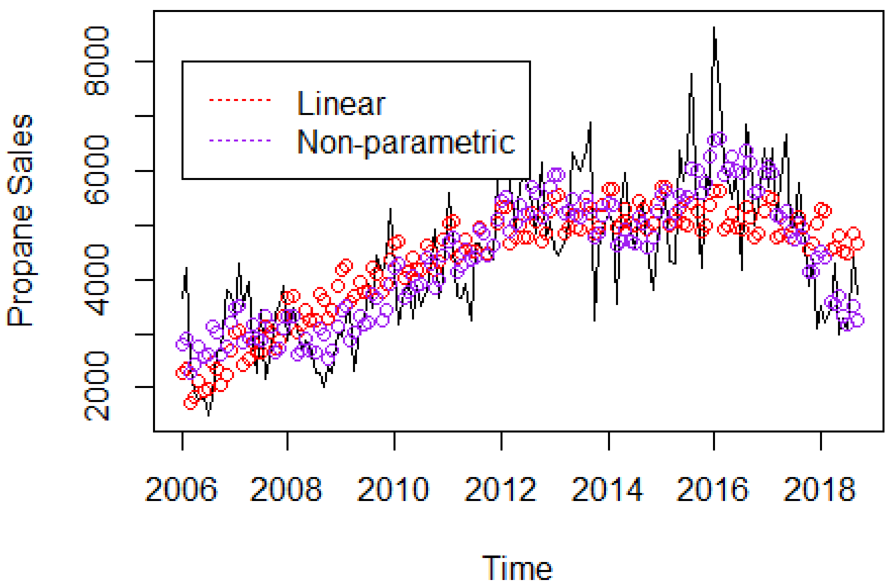
For this question, I attempted to estimate both the trend and seasonality in sales for the three products using different models. For each product, I fit one model that used linear regression for the trend and seasonality. I also fit a second model that used non-parametric trend and linear seasonality. I then plot the fitted values vs. originals along with the residuals and ACF of the residuals for each model and product.



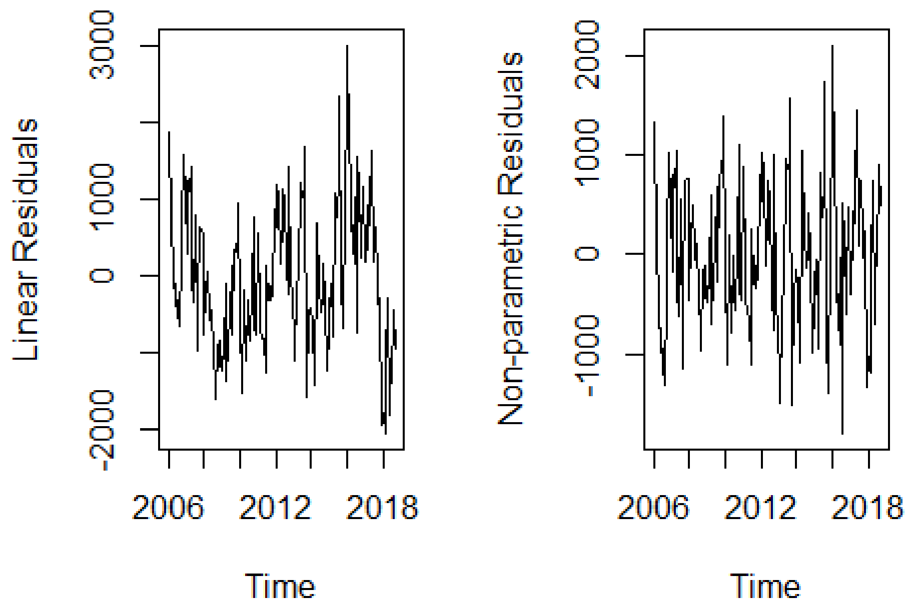
No. 2 Sales
25000
20000
15000
Linear
Non-parametric
2006
2008
2010
2012
2014
2016
2018
Time
Linear Residuals
3000
1000
0
-2000
2006
2012
2018
Time
Non-parametric Residuals
1000
0
-2000
2006
2012
2018
Time

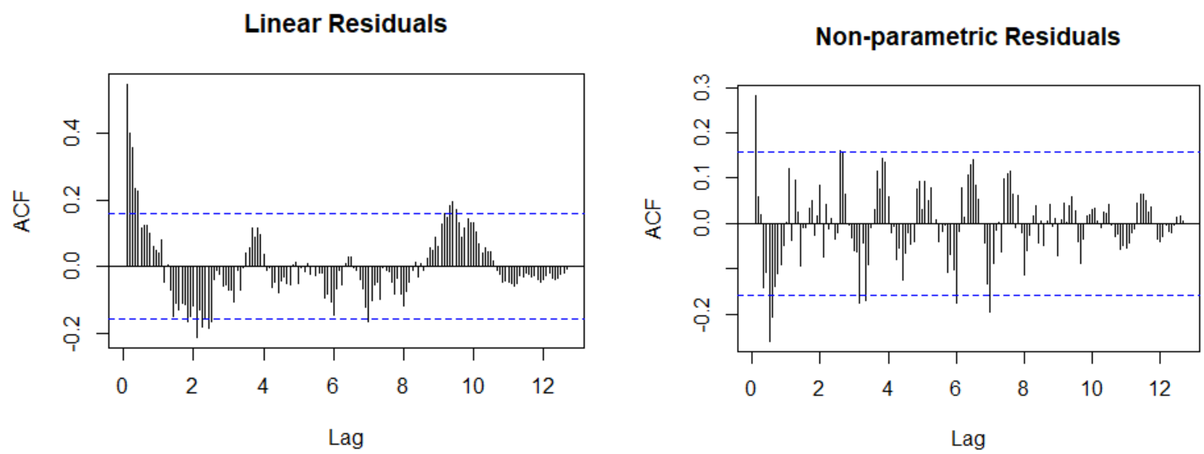


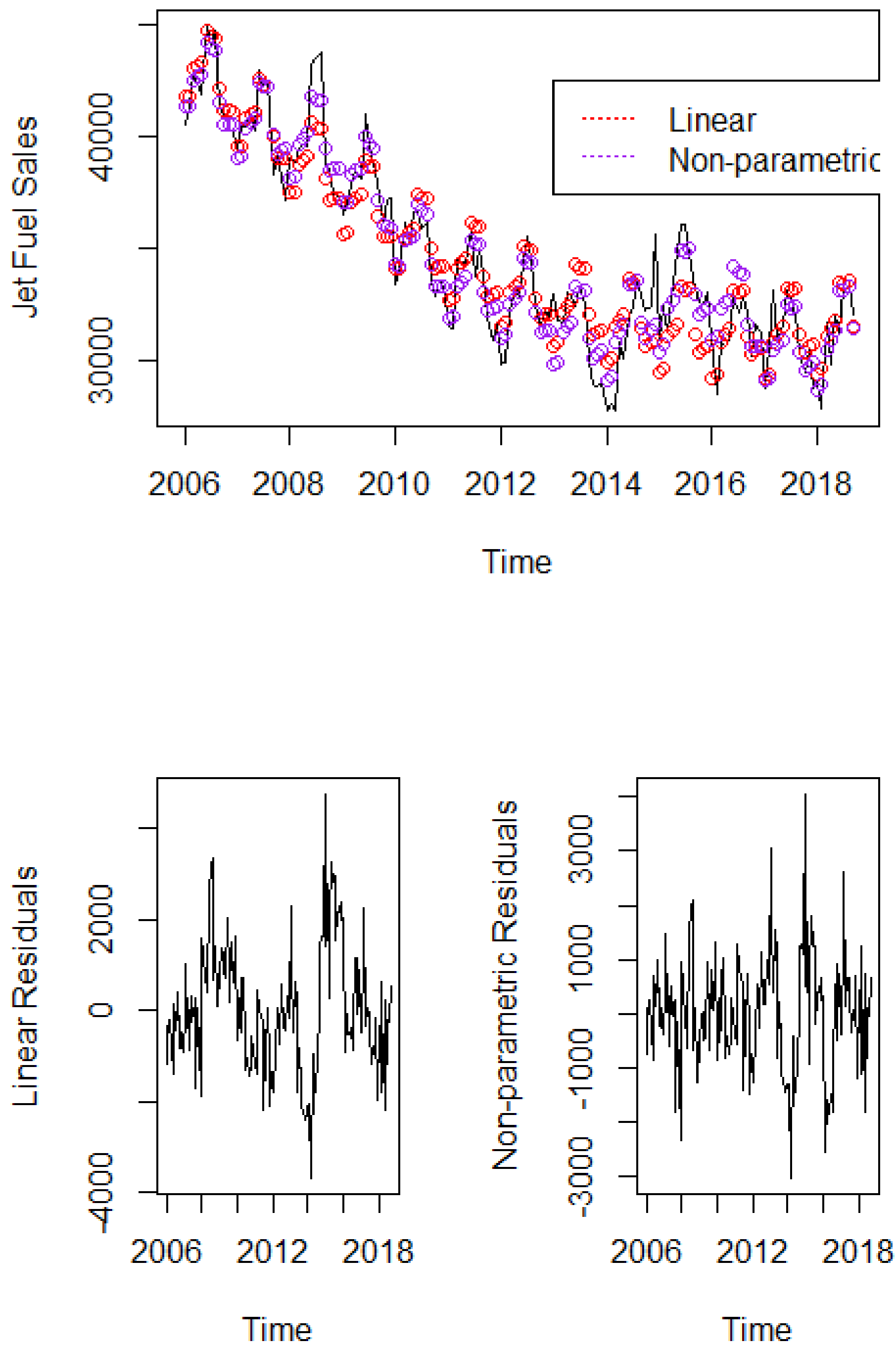
Linear Residuals
Non-parametric Residuals
ACF
Lag



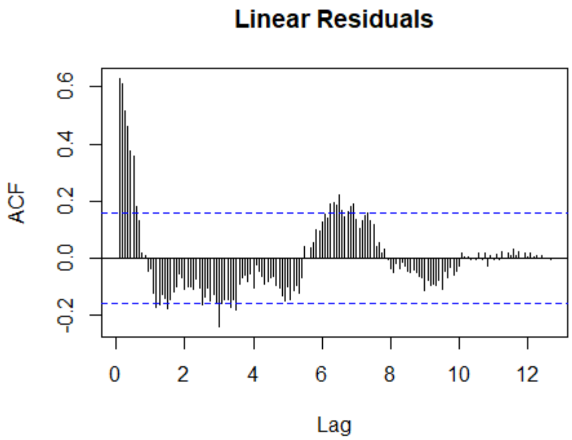
Linear
Non-parametric
Propane Sales
Time

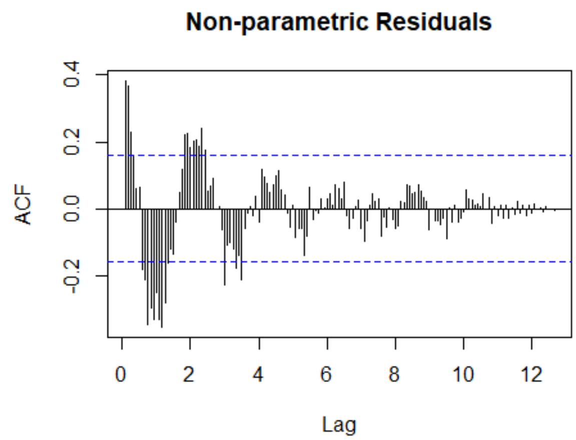

Linear Residuals
3000
1000
0
-2000
2006
2012
2018
Time
Non-parametric Residuals
2000
1000
0
-1000
2006
2012
2018
Time


Linear Residuals
ACF
0.4
0.2
0.0
-0.2
0
2
4
6
8
10
12
Lag
Non-parametric Residuals
ACF
0.3
0.2
0.1
0.0
-0.2
0
2
4
6
8
10
12
Lag



Linear
Non-parametric
Jet Fuel Sales
30000
40000
2006
2008
2010
2012
2014
2016
2018
Time
Linear Residuals
-4000
0
2000
2006
2012
2018
Time
Non-parametric Residuals
-3000
-1000
1000
3000
2006
2012
2018
Time





Finally, I put everything together and fit two models combining trend and seasonality estimation for all three products (No. 2, Propane, and Jet Fuel). Looking at the plots of the model fits, these models seem to closely follow the original data. They capture both the trend and an element of seasonality. Both models fit well for No. 2. For propane, the non-parametric model seems to have fit the data slightly better. For Jet Fuel, both models again fit well.

Looking at the residuals and ACF plots, we see some interesting observations. Most of the models seem to be roughly centered around zero, in line with the constant mean assumption of stationarity. None of the models seem to have large spikes in the residuals, which would violate the finite variance assumption. However, looking at the ACF plots reveals that there still seems to be non-constant variance for many of these models. Specifically, No. 2 and Jet Fuel seem to still have some kind of cyclical trend. Propane's ACF plots also show variance that decreases over time. It could be that there is another cause of a cyclical trend that is not captured here. One example could be a higher demand for jet fuel around the holidays by the airlines.

Based on these results, I would not conclude that the residuals satisfy all the assumptions of stationarity yet. Additional analysis would have to be done.

Appendix: R Code

```
#Read Data
data=read.csv(file.choose(),header=TRUE)
no2=as.vector(data[,2])
propane=data[,3]
jet=data[,4]

no2 = ts(no2, start=2006, frequency = 12)
propane = ts(propane, start=2006, frequency = 12)
jet = ts(jet, start=2006, frequency = 12)

ts.plot(no2, ylab="No. 2 Sales")
ts.plot(propane, ylab="Propane Sales")
ts.plot(jet, ylab="Jet Fuel Sales")

diff.no2 = diff(no2,differences=1)
diff.propane = diff(propane,differences=1)
diff.jet = diff(jet,differences=1)

ts.plot(diff.no2, ylab="No. 2 Diff")
ts.plot(diff.propane, ylab="Propane Diff")
ts.plot(diff.jet, ylab="Jet Fuel Diff")

##Trend Estimation
#No. 2
time.pts = c(1:nrow(data))
time.pts = c(time.pts - min(time.pts))/max(time.pts)
#Moving Average
mav.fit.no2 = ksmooth(time.pts, no2, kernel = "box")
no2.fit.mav = ts(mav.fit.no2$y, start = 2006, frequency = 12)
#Parametric Quadratic Polynomial
x1 = time.pts
x2 = time.pts^2
lm.fit.no2 = lm(no2~x1+x2)
no2.fit.lm = ts(fitted(lm.fit.no2),start=2006,frequency=12)
#Local Polynomial
loc.fit.no2 = loess(no2~time.pts)
no2.fit.loc = ts(fitted(loc.fit.no2),start=2006,frequency=12)
#Splines
library(mgcv)
gam.fit.no2 = gam(no2~s(time.pts))
no2.fit.gam = ts(fitted(gam.fit.no2),start=2006,frequency=12)
#Plot Fitted Values
ts.plot(no2, ylab="No. 2 Sales")
points(no2.fit.mav, col="red")
points(no2.fit.lm, col="purple")
points(no2.fit.loc, col="blue")
points(no2.fit.gam, col="brown")
legend(x=2016,y=25000,legend=c("MAV","LM","LOESS","GAM"),lty = 3, col=c("red"
```

```r
,"purple","blue","brown"))
#Plot Residuals
par(mfrow=c(2,2))
ts.plot(no2-no2.fit.mav, ylab="MAV Residuals")
ts.plot(no2-no2.fit.lm, ylab="LM Residuals")
ts.plot(no2-no2.fit.loc, ylab="LOESS Residuals")
ts.plot(no2-no2.fit.gam, ylab="GAM Residuals")
#Plot ACF of Residuals
acf(no2-no2.fit.mav,lag.max=12*13,main="MAV Residuals")
acf(no2-no2.fit.lm, lag.max=12*13,main="LM Residuals")
acf(no2-no2.fit.loc,lag.max=12*13,main="LOESS Residuals")
acf(no2-no2.fit.gam, lag.max=12*13, main="GAM Residuals")
#Propane
#Moving Average
mav.fit.propane = ksmooth(time.pts, propane, kernel = "box")
propane.fit.mav = ts(mav.fit.propane$y, start = 2006, frequency = 12)
#Parametric Quadratic Polynomial
lm.fit.propane = lm(propane~x1+x2)
propane.fit.lm = ts(fitted(lm.fit.propane),start=2006,frequency=12)
#Local Polynomial
loc.fit.propane = loess(propane~time.pts)
propane.fit.loc = ts(fitted(loc.fit.propane),start=2006,frequency=12)
#Splines
gam.fit.propane = gam(propane~s(time.pts))
propane.fit.gam = ts(fitted(gam.fit.propane),start=2006,frequency=12)
#Plot Fitted Values
par(mfrow=c(1,1))
ts.plot(propane, ylab="Propane Sales")
points(propane.fit.mav, col="red")
points(propane.fit.lm, col="purple")
points(propane.fit.loc, col="blue")
points(propane.fit.gam, col="brown")
legend(x=2006,y=8000,legend=c("MAV","LM","LOESS","GAM"),lty = 3, col=c("red",
"purple","blue","brown"))
#Plot Residuals
par(mfrow=c(2,2))
ts.plot(propane-propane.fit.mav, ylab="MAV Residuals")
ts.plot(propane-propane.fit.lm, ylab="LM Residuals")
ts.plot(propane-propane.fit.loc, ylab="LOESS Residuals")
ts.plot(propane-propane.fit.gam, ylab="GAM Residuals")
#Plot ACF of Residuals
acf(propane-propane.fit.mav,lag.max=12*13,main="MAV Residuals")
acf(propane-propane.fit.lm, lag.max=12*13,main="LM Residuals")
acf(propane-propane.fit.loc,lag.max=12*13,main="LOESS Residuals")
acf(propane-propane.fit.gam, lag.max=12*13, main="GAM Residuals")
#Jet Fuel
#Moving Average
mav.fit.jet = ksmooth(time.pts, jet, kernel = "box")
jet.fit.mav = ts(mav.fit.jet$y, start = 2006, frequency = 12)
#Parametric Quadratic Polynomial
```

```r
lm.fit.jet = lm(jet~x1+x2)
jet.fit.lm = ts(fitted(lm.fit.jet),start=2006,frequency=12)
#Local Polynomial
loc.fit.jet = loess(jet~time.pts)
jet.fit.loc = ts(fitted(loc.fit.jet),start=2006,frequency=12)
#Splines
gam.fit.jet = gam(jet~s(time.pts))
jet.fit.gam = ts(fitted(gam.fit.jet),start=2006,frequency=12)
#Plot Fitted Values
par(mfrow=c(1,1))
ts.plot(jet, ylab="Jet Fuel Sales")
points(jet.fit.mav, col="red")
points(jet.fit.lm, col="purple")
points(jet.fit.loc, col="blue")
points(jet.fit.gam, col="brown")
legend(x=2015,y=42500,legend=c("MAV","LM","LOESS","GAM"),lty = 3, col=c("red"
,"purple","blue","brown"))
#Plot Residuals
par(mfrow=c(2,2))
ts.plot(jet-jet.fit.mav, ylab="MAV Residuals")
ts.plot(jet-jet.fit.lm, ylab="LM Residuals")
ts.plot(jet-jet.fit.loc, ylab="LOESS Residuals")
ts.plot(jet-jet.fit.gam, ylab="GAM Residuals")
#Plot ACF of Residuals
acf(jet-jet.fit.mav,lag.max=12*13,main="MAV Residuals")
acf(jet-jet.fit.lm, lag.max=12*13,main="LM Residuals")
acf(jet-jet.fit.loc,lag.max=12*13,main="LOESS Residuals")
acf(jet-jet.fit.gam, lag.max=12*13, main="GAM Residuals")

##Seasonality
#No. 2
#ANOVA
library(TSA)
par(mfrow=c(1,1))
month = season(no2)
anova.fit.no2 = lm(no2~month-1)
no2.fit.anova = ts(fitted(anova.fit.no2), start=2006, frequency=12)
#Cos-Sin
har.no2 = harmonic(no2,1)
har.fit.no2 = lm(no2~har.no2)
no2.fit.har = ts(fitted(har.fit.no2), start=2006, frequency=12)
#Plot Fitted Values
ts.plot(no2, ylab="No. 2 Sales")
points(no2.fit.anova, col="red")
points(no2.fit.har, col="purple")
legend(x=2015,y=25000,legend=c("ANOVA","Cos-Sin"),lty = 3, col=c("red","purpl
e"))
#Plot Residuals
par(mfrow=c(1,2))
ts.plot(no2-no2.fit.anova, ylab="ANOVA Residuals")
```

```r
ts.plot(no2-no2.fit.har, ylab="Cos-Sin Residuals")
#Plot ACF of Residuals
par(mfrow=c(2,1))
acf(no2-no2.fit.anova,lag.max=12*13,main="ANOVA Residuals")
acf(no2-no2.fit.har, lag.max=12*13,main="Cos-Sin Residuals")
#Propane
#ANOVA
par(mfrow=c(1,1))
anova.fit.propane = lm(propane~month-1)
propane.fit.anova = ts(fitted(anova.fit.propane), start=2006, frequency=12)
#Cos-Sin
har.propane = harmonic(propane,1)
har.fit.propane = lm(propane~har.propane)
propane.fit.har = ts(fitted(har.fit.propane), start=2006, frequency=12)
#Plot Fitted Values
ts.plot(propane, ylab="Propane Sales")
points(propane.fit.anova, col="red")
points(propane.fit.har, col="purple")
legend(x=2006,y=8000,legend=c("ANOVA","Cos-Sin"),lty = 3, col=c("red","purple
"))
#Plot Residuals
par(mfrow=c(1,2))
ts.plot(propane-propane.fit.anova, ylab="ANOVA Residuals")
ts.plot(propane-propane.fit.har, ylab="Cos-Sin Residuals")
#Plot ACF of Residuals
par(mfrow=c(2,1))
acf(propane-propane.fit.anova,lag.max=12*13,main="ANOVA Residuals")
acf(propane-propane.fit.har, lag.max=12*13,main="Cos-Sin Residuals")
#Jet Fuel
#ANOVA
par(mfrow=c(1,1))
anova.fit.jet = lm(jet~month-1)
jet.fit.anova = ts(fitted(anova.fit.jet), start=2006, frequency=12)
#Cos-Sin
har.jet = harmonic(jet,1)
har.fit.jet = lm(jet~har.jet)
jet.fit.har = ts(fitted(har.fit.jet), start=2006, frequency=12)
#Plot Fitted Values
ts.plot(jet, ylab="Jet Fuel Sales")
points(jet.fit.anova, col="red")
points(jet.fit.har, col="purple")
legend(x=2015,y=42500,legend=c("ANOVA","Cos-Sin"),lty = 3, col=c("red","purpl
e"))
#Plot Residuals
par(mfrow=c(1,2))
ts.plot(jet-jet.fit.anova, ylab="ANOVA Residuals")
ts.plot(jet-jet.fit.har, ylab="Cos-Sin Residuals")
#Plot ACF of Residuals
par(mfrow=c(2,1))
acf(jet-jet.fit.anova,lag.max=12*13,main="ANOVA Residuals")
```

```r
acf(jet-jet.fit.har, lag.max=12*13,main="Cos-Sin Residuals")

##Trend and Seasonality Estimation
#No. 2
#Linear
linear.fit.no2 = lm(no2~x1+x2+month-1)
no2.fit.linear = ts(fitted(linear.fit.no2), start=2006, frequency=12)
#Non-parametric
nonpar.fit.no2 = gam(no2~s(time.pts)+month-1)
no2.fit.nonpar = ts(fitted(nonpar.fit.no2), start=2006, frequency=12)
#Plot Fitted Values
par(mfrow=c(1,1))
ts.plot(no2, ylab="No. 2 Sales")
points(no2.fit.linear, col="red")
points(no2.fit.nonpar, col="purple")
legend(x=2014,y=25000,legend=c("Linear","Non-parametric"),lty = 3, col=c("red
","purple"))
#Plot Residuals
par(mfrow=c(1,2))
ts.plot(no2-no2.fit.linear, ylab="Linear Residuals")
ts.plot(no2-no2.fit.nonpar, ylab="Non-parametric Residuals")
#Plot ACF of Residuals
par(mfrow=c(2,1))
acf(no2-no2.fit.linear,lag.max=12*13,main="Linear Residuals")
acf(no2-no2.fit.nonpar, lag.max=12*13,main="Non-parametric Residuals")
#Propane
#Linear
linear.fit.propane = lm(propane~x1+x2+month-1)
propane.fit.linear = ts(fitted(linear.fit.propane), start=2006, frequency=12)
#Non-parametric
nonpar.fit.propane = gam(propane~s(time.pts)+month-1)
propane.fit.nonpar = ts(fitted(nonpar.fit.propane), start=2006, frequency=12)
#Plot Fitted Values
par(mfrow=c(1,1))
ts.plot(propane, ylab="Propane Sales")
points(propane.fit.linear, col="red")
points(propane.fit.nonpar, col="purple")
legend(x=2006,y=8000,legend=c("Linear","Non-parametric"),lty = 3, col=c("red"
,"purple"))
#Plot Residuals
par(mfrow=c(1,2))
ts.plot(propane-propane.fit.linear, ylab="Linear Residuals")
ts.plot(propane-propane.fit.nonpar, ylab="Non-parametric Residuals")
#Plot ACF of Residuals
par(mfrow=c(2,1))
acf(propane-propane.fit.linear,lag.max=12*13,main="Linear Residuals")
acf(propane-propane.fit.nonpar, lag.max=12*13,main="Non-parametric Residuals"
)
#Jet Fuel
#Linear
```

```
linear.fit.jet = lm(jet~x1+x2+month-1)
jet.fit.linear = ts(fitted(linear.fit.jet), start=2006, frequency=12)
#Non-parametric
nonpar.fit.jet = gam(jet~s(time.pts)+month-1)
jet.fit.nonpar = ts(fitted(nonpar.fit.jet), start=2006, frequency=12)
#Plot Fitted Values
par(mfrow=c(1,1))
ts.plot(jet, ylab="Jet Fuel Sales")
points(jet.fit.linear, col="red")
points(jet.fit.nonpar, col="purple")
legend(x=2015,y=42500,legend=c("Linear","Non-parametric"),lty = 3, col=c("red
","purple"))
#Plot Residuals
par(mfrow=c(1,2))
ts.plot(jet-jet.fit.linear, ylab="Linear Residuals")
ts.plot(jet-jet.fit.nonpar, ylab="Non-parametric Residuals")
#Plot ACF of Residuals
par(mfrow=c(2,1))
acf(jet-jet.fit.linear,lag.max=12*13,main="Linear Residuals")
acf(jet-jet.fit.nonpar, lag.max=12*13,main="Non-parametric Residuals")
```