Economics 403B: Project 1

Ross Lewis, Adam Jacobson, David Contento

January 16, 2019

TODOS discuss - adam / all of us introduction - david conclusions - ross code fixes predition intervals - david

I

Introduction

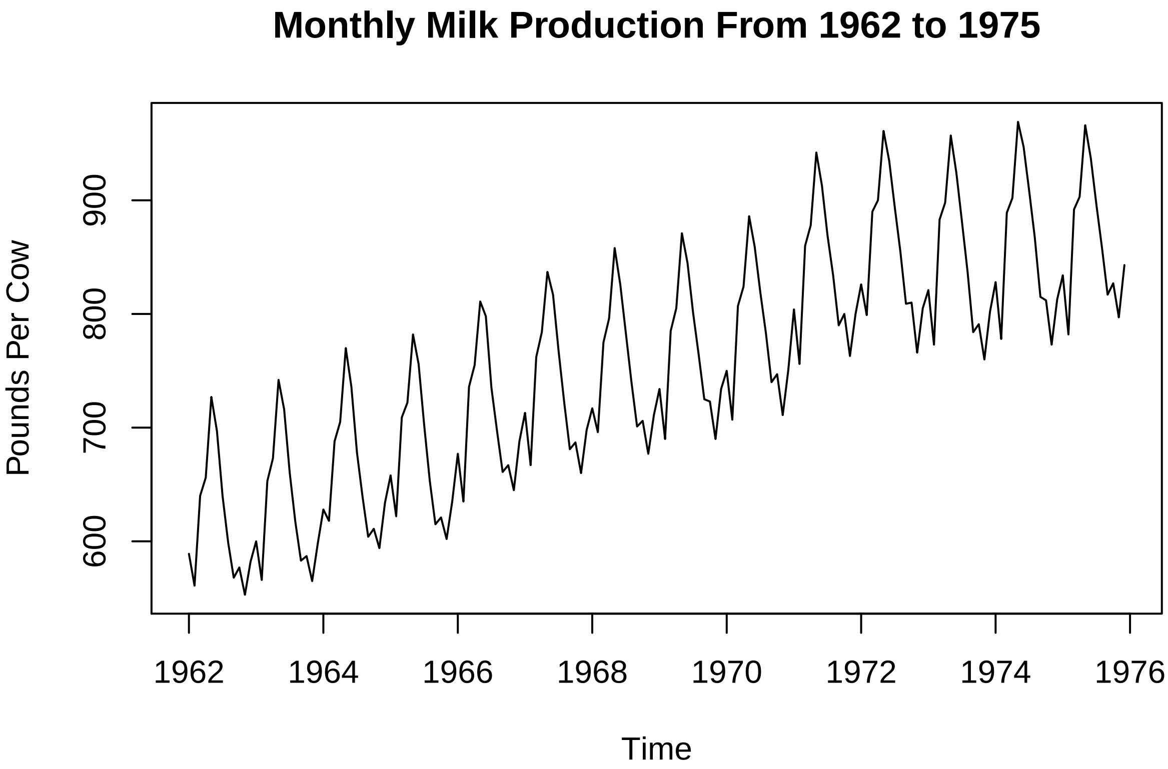
TODO

II

Results

1a

```
setwd('C:/Users/rossw/Documents/MAE Program/Q2/Applied Econometrics 403B/Project 1')
data = read.csv("monthly-milk-production-pounds-p.csv",header = F)
names(data) = c('date','milkproduction')
data = na.exclude(data)
attach(data)
datats = ts(milkproduction,start=1962,freq=12)
#datats = na.exclude(datats)
time = seq(1962,1975.91666666666,length=length(datats))
plot.ts(datats,main='Monthly Milk Production From 1962 to 1975',ylab='Pounds Per Cow')
```



```
#datats
#time
#na.exclude(datats)
```

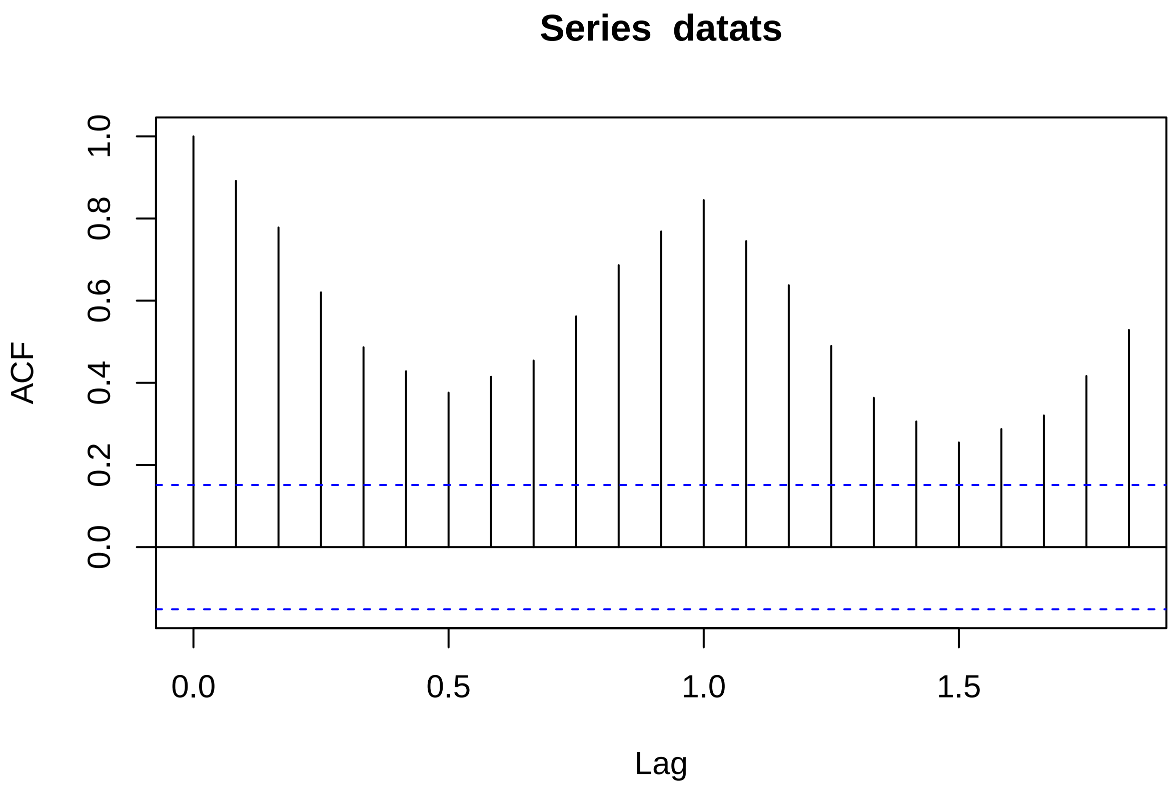
1b

TODO

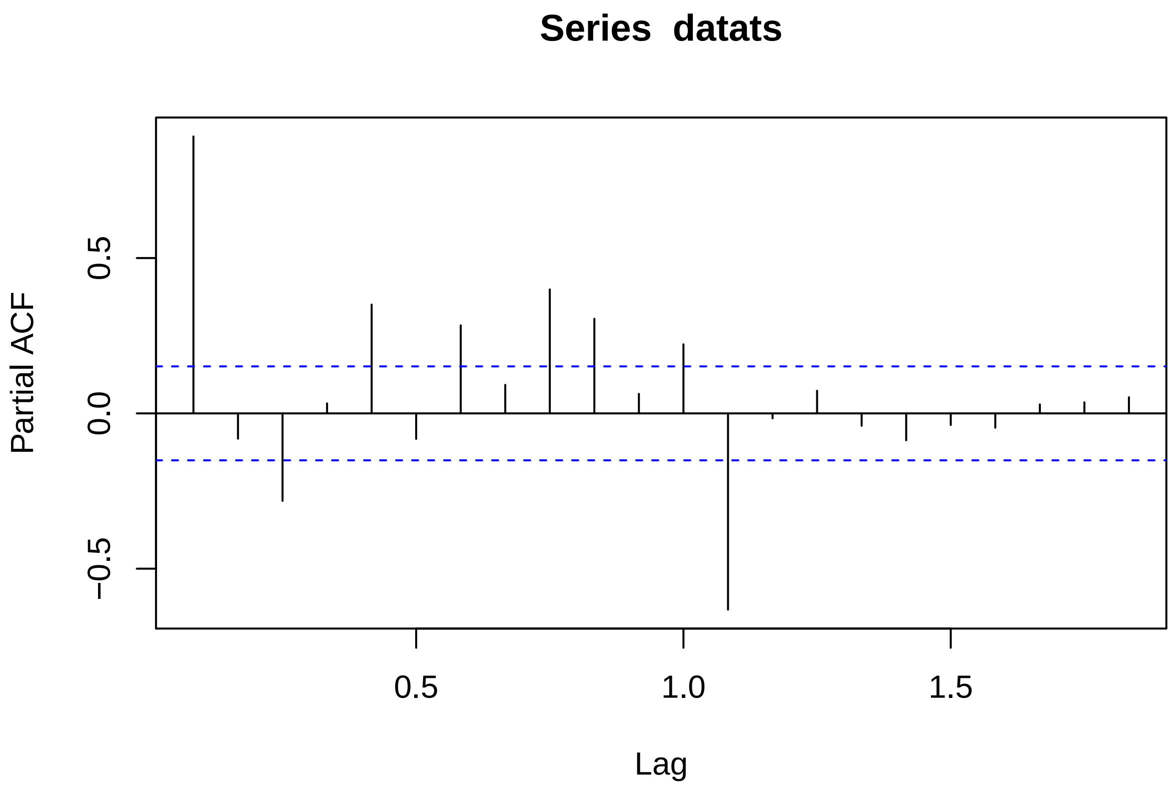
1c

We observe large amounts of autocorrelation obviously. This shows nonstationary.

```
#acf and pcf plots
acf(datats)
```

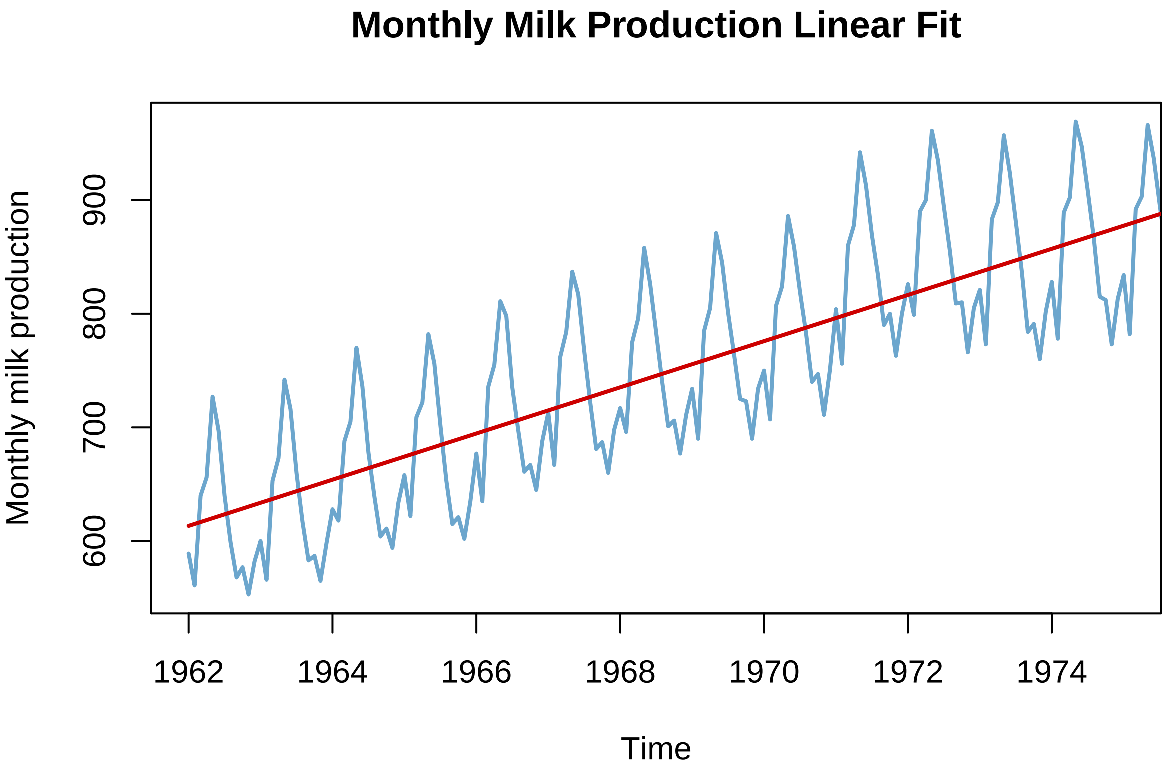


```
pacf(datats)
```

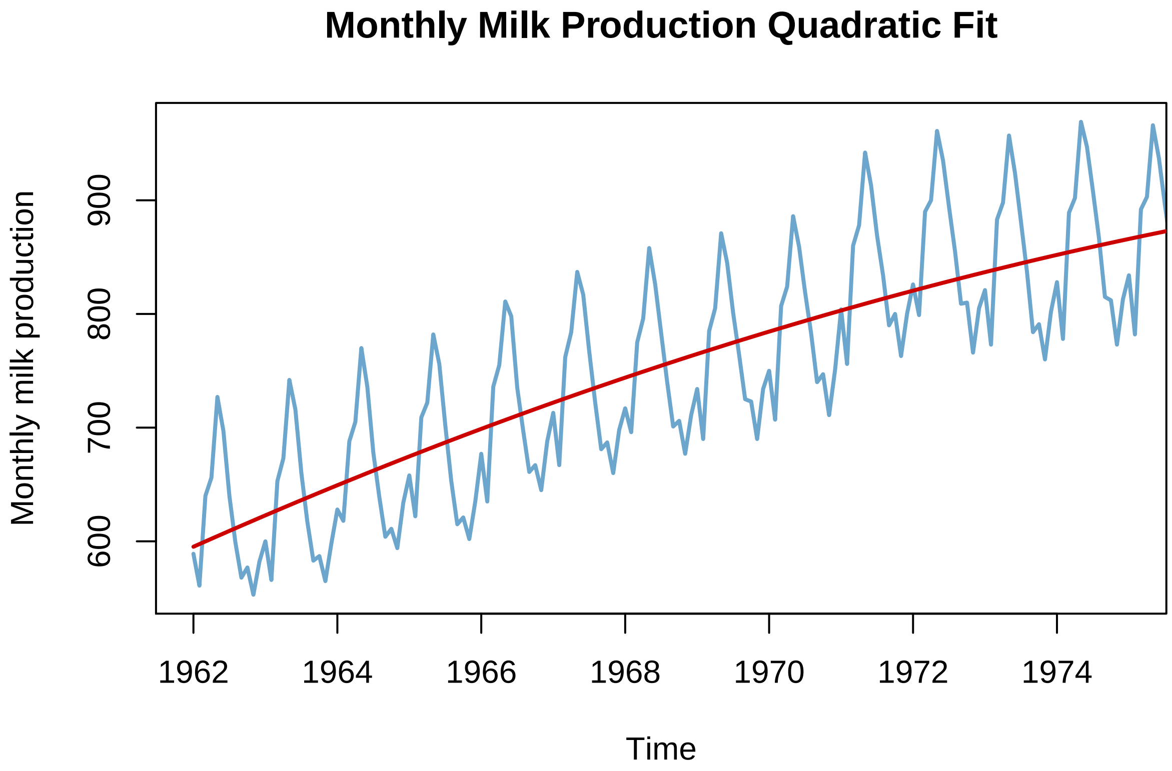


1d

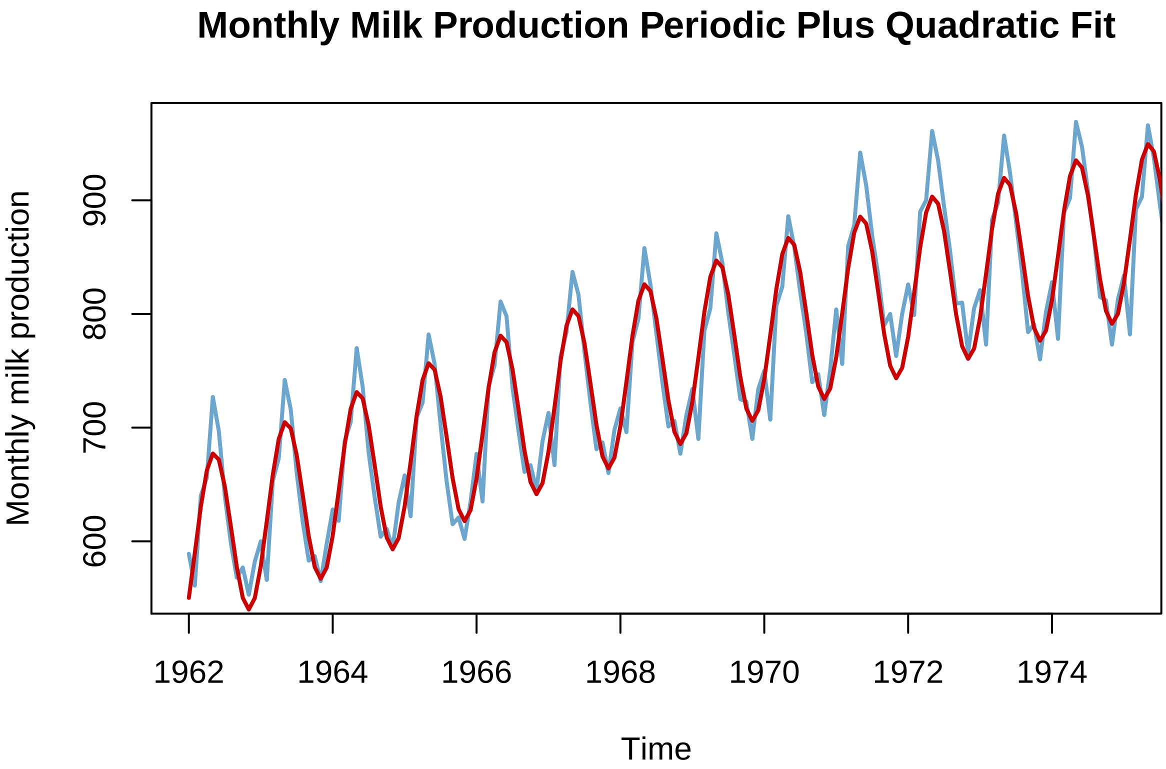
```
#, fig.width=6, fig.height=6
#Linear Fit
mod1=lm(datats~time)
#par(mfrow=c(2,1))
plot(datats, ylab="Monthly milk production",main='Monthly Milk Production Linear Fit', xlab="Time", lwd=
#plot(datats)
lines(time,mod1$fitted.values,col="red3",lwd=2)
```



```
#quadratic fit
mod2=lm(datats~time+I(time^2))
#par(mfrow=c(2,1))
plot(datats,ylab="Monthly milk production",main='Monthly Milk Production Quadratic Fit', xlab="Time", lw
lines(time,mod2$fitted.values,col="red3",lwd=2)
```

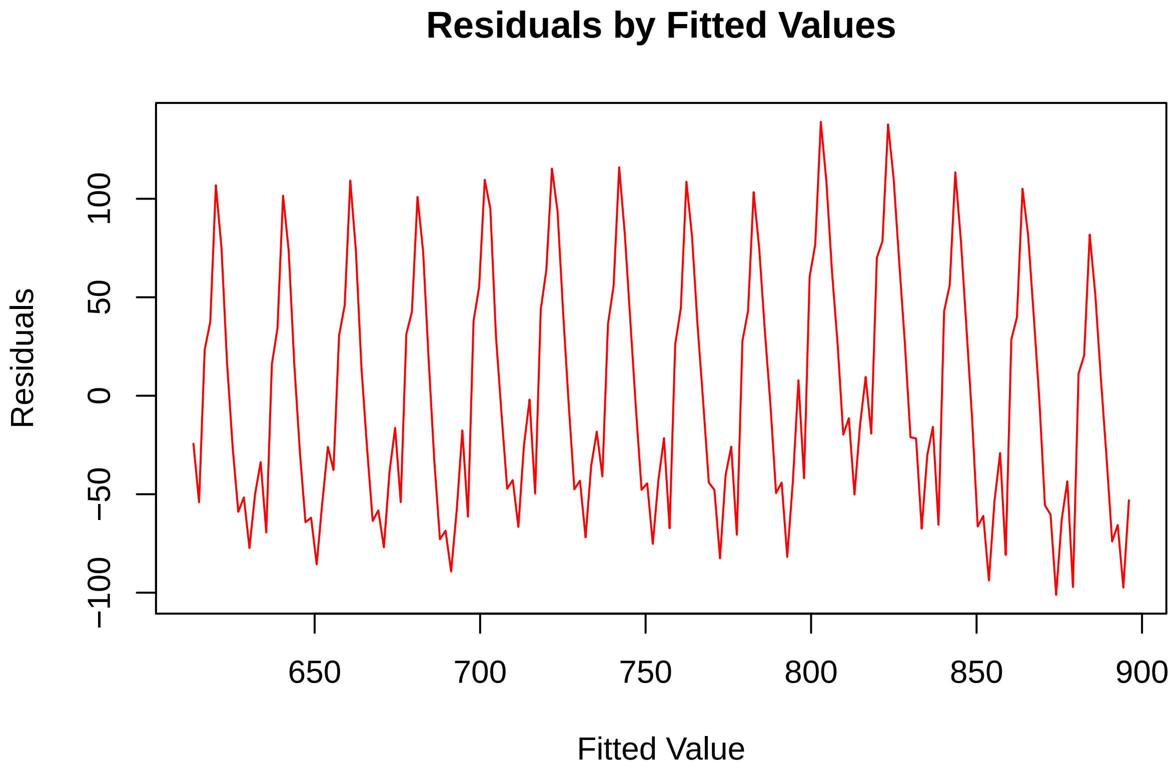


```
#periodic + quadratic fit
sin.t<-sin(2*pi*time)
cos.t<-cos(2*pi*time)
mod3=lm(datats~time+I(time^2) + sin.t + cos.t)
#par(mfrow=c(2,1))
plot(datats,ylab="Monthly milk production",main='Monthly Milk Production Periodic Plus Quadratic Fit', 
lines(time,mod3$fitted.values,col="red3",lwd=2)
```

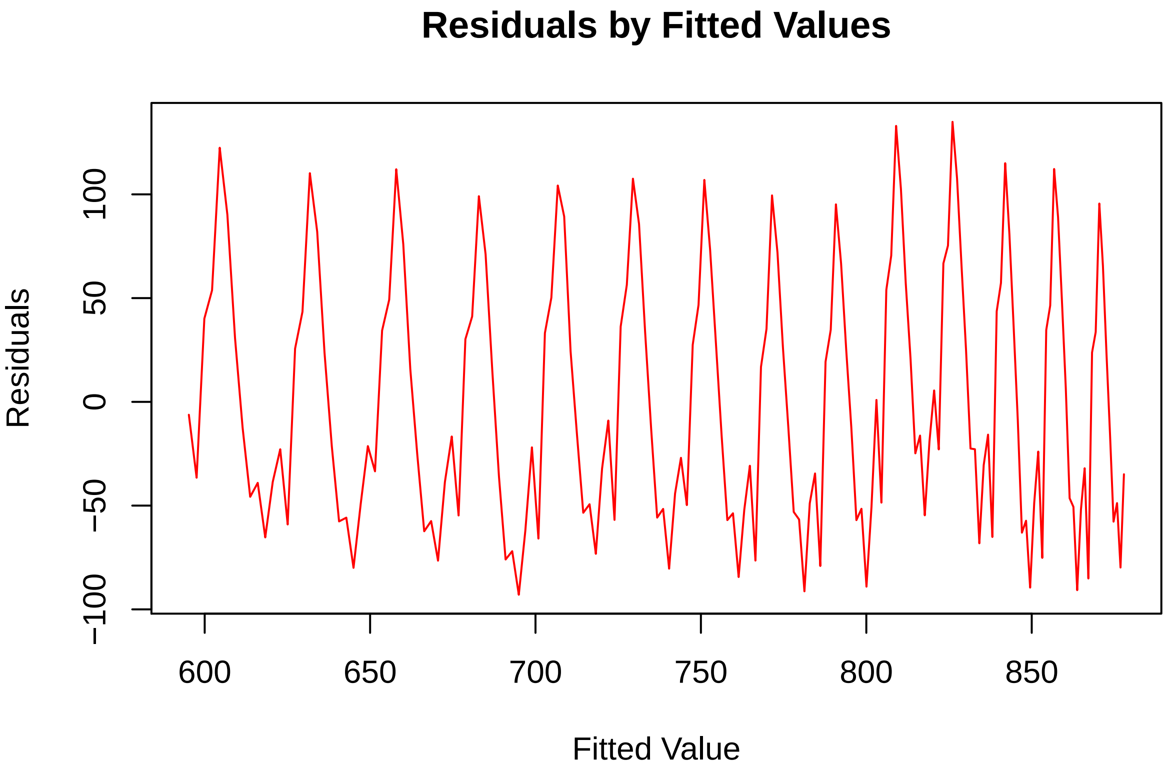


1e

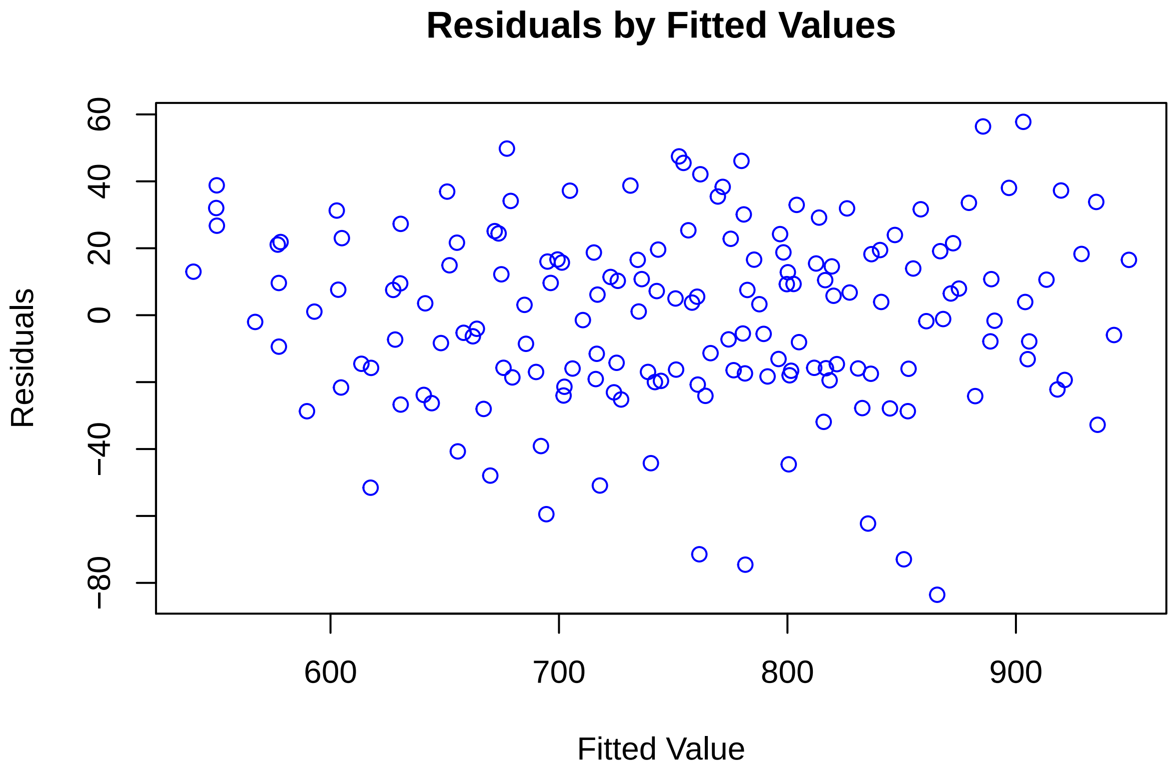
```
#plot of residuals for linear
plot(mod1$fitted.values,mod1$residuals, main='Residuals by Fitted Values', ylab="Residuals",type='l',xla
```



```
plot(mod2$fitted.values,mod2$residuals, main='Residuals by Fitted Values', ylab="Residuals",type='l',xla
```



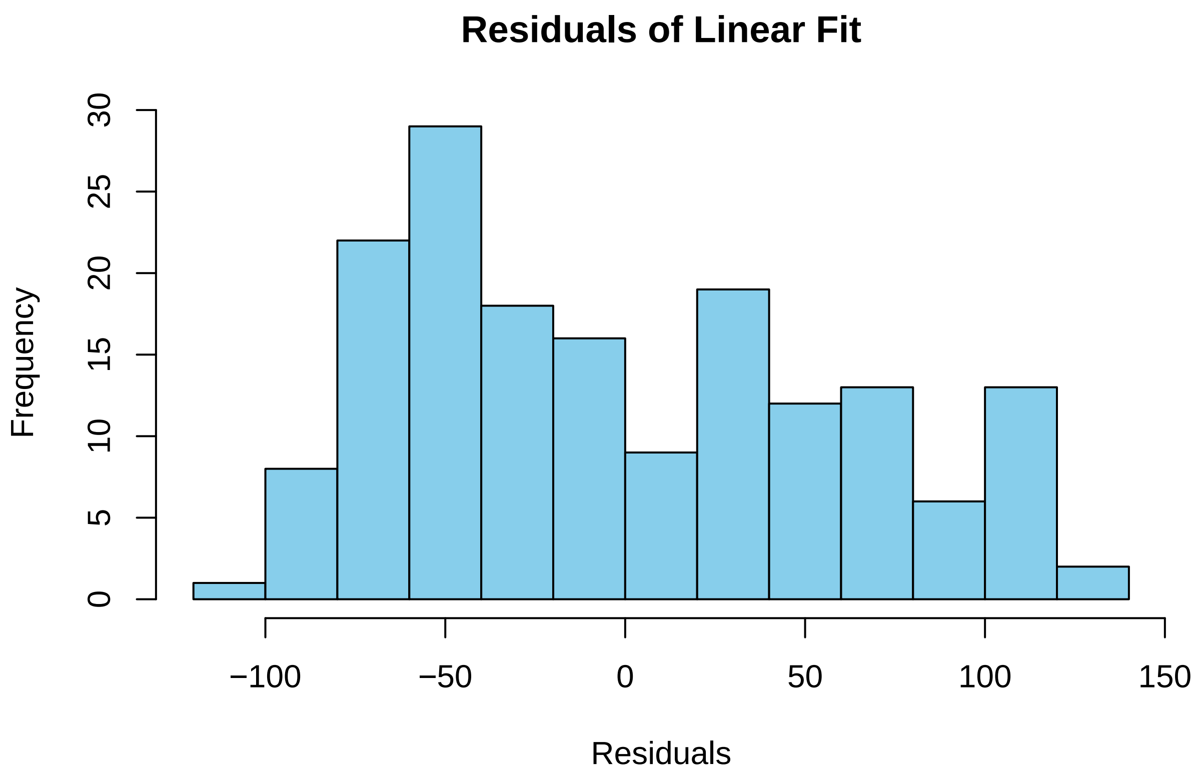
```
plot(mod3$fitted.values,mod3$residuals, main='Residuals by Fitted Values', ylab="Residuals",xlab="Fitte
```



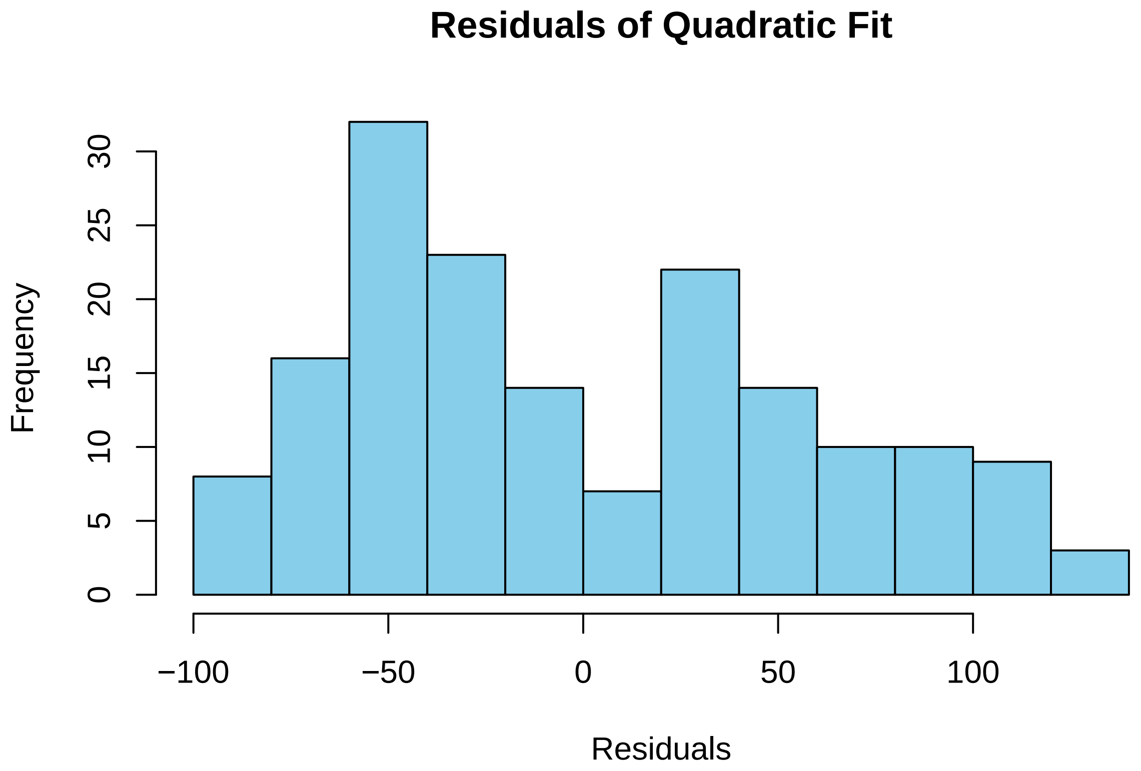
1f

TODO discuss

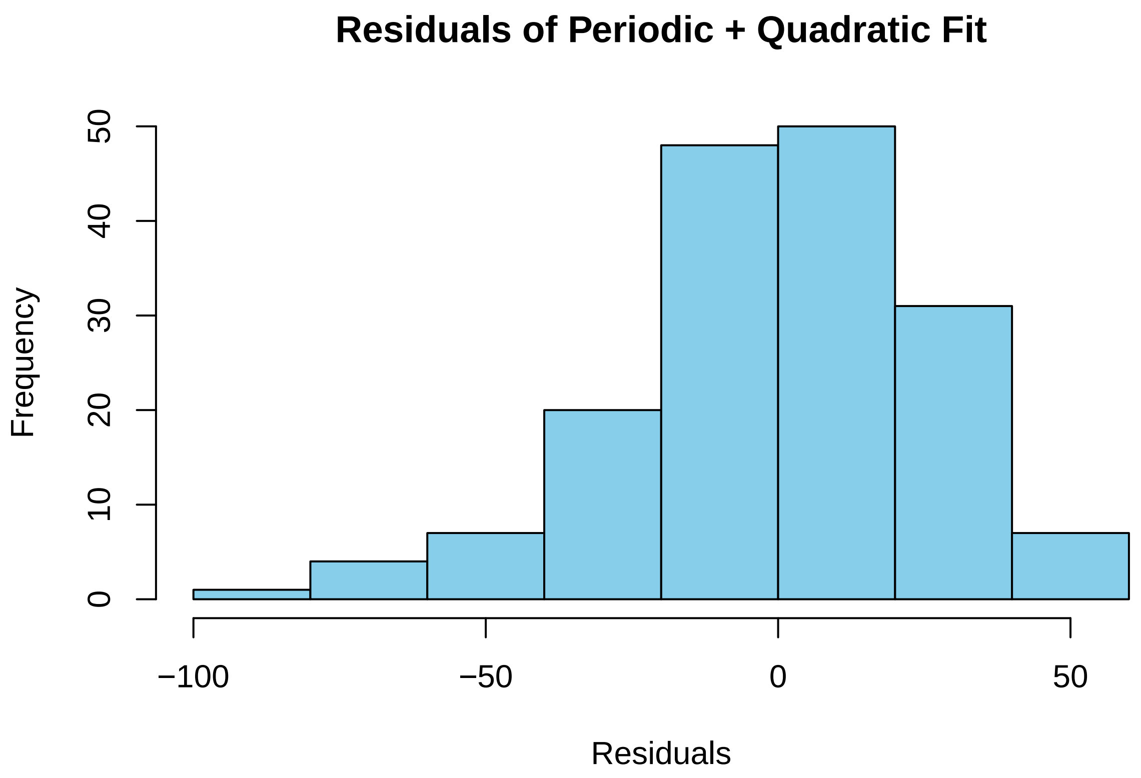
```
hist(mod1$residuals,col='skyblue',xlab='Residuals',main='Residuals of Linear Fit')
```



```r
hist(mod2$residuals,col='skyblue',xlab='Residuals',main='Residuals of Quadratic Fit')
```



```
hist(mod3$residuals,col='skyblue',xlab='Residuals',main='Residuals of Periodic + Quadratic Fit')
```



1g

TODO DISCUSS

```
library(tsoutliers)
JarqueBera.test(mod1$residuals)
```

```
##
##  Jarque Bera Test
##
## data:  mod1$residuals
## X-squared = 10.77, df = 2, p-value = 0.004585
##
##
##  Skewness
##
## data:  mod1$residuals
## statistic = 0.40457, p-value = 0.03229
##
##
##  Kurtosis
##
## data:  mod1$residuals
## statistic = 2.0599, p-value = 0.01287
```

```
JarqueBera.test(mod2$residuals)
```

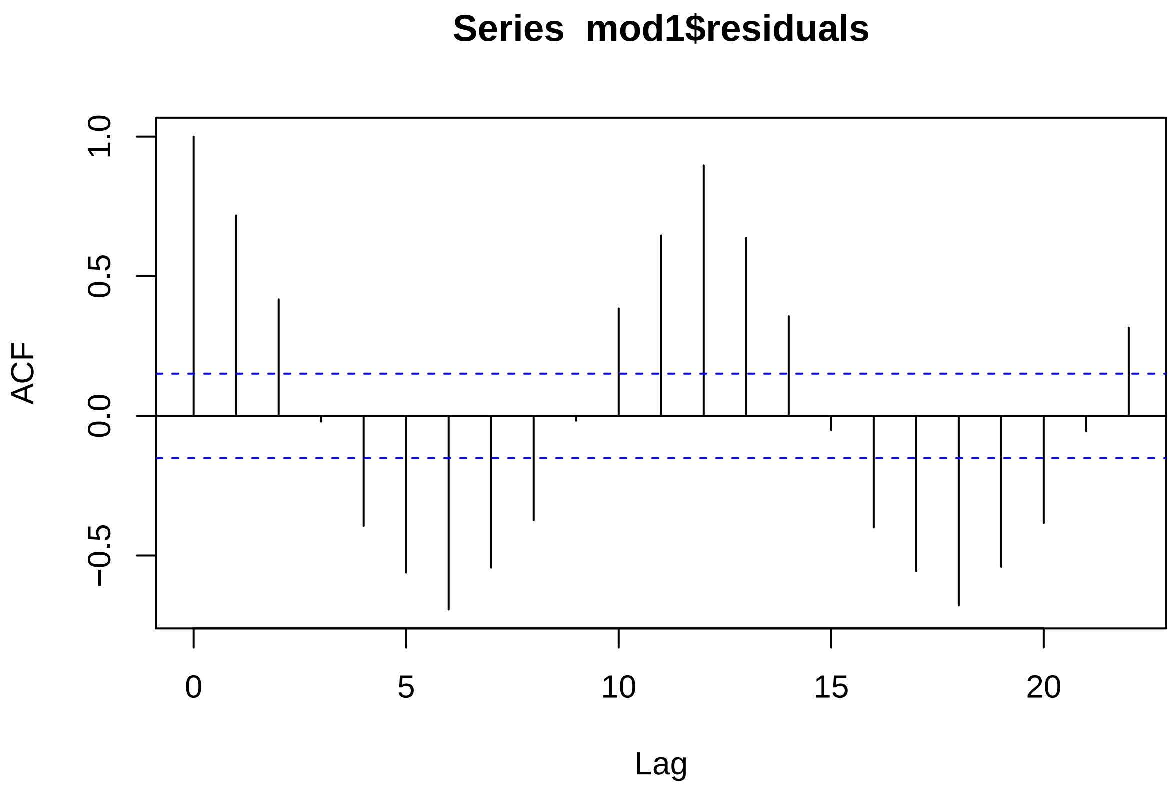
```
##
##  Jarque Bera Test
##
## data:  mod2$residuals
## X-squared = 11.24, df = 2, p-value = 0.003624
##
##
##  Skewness
##
## data:  mod2$residuals
## statistic = 0.41165, p-value = 0.02939
##
##
##  Kurtosis
##
## data:  mod2$residuals
## statistic = 2.0367, p-value = 0.01081
```

```
JarqueBera.test(mod3$residuals)
```

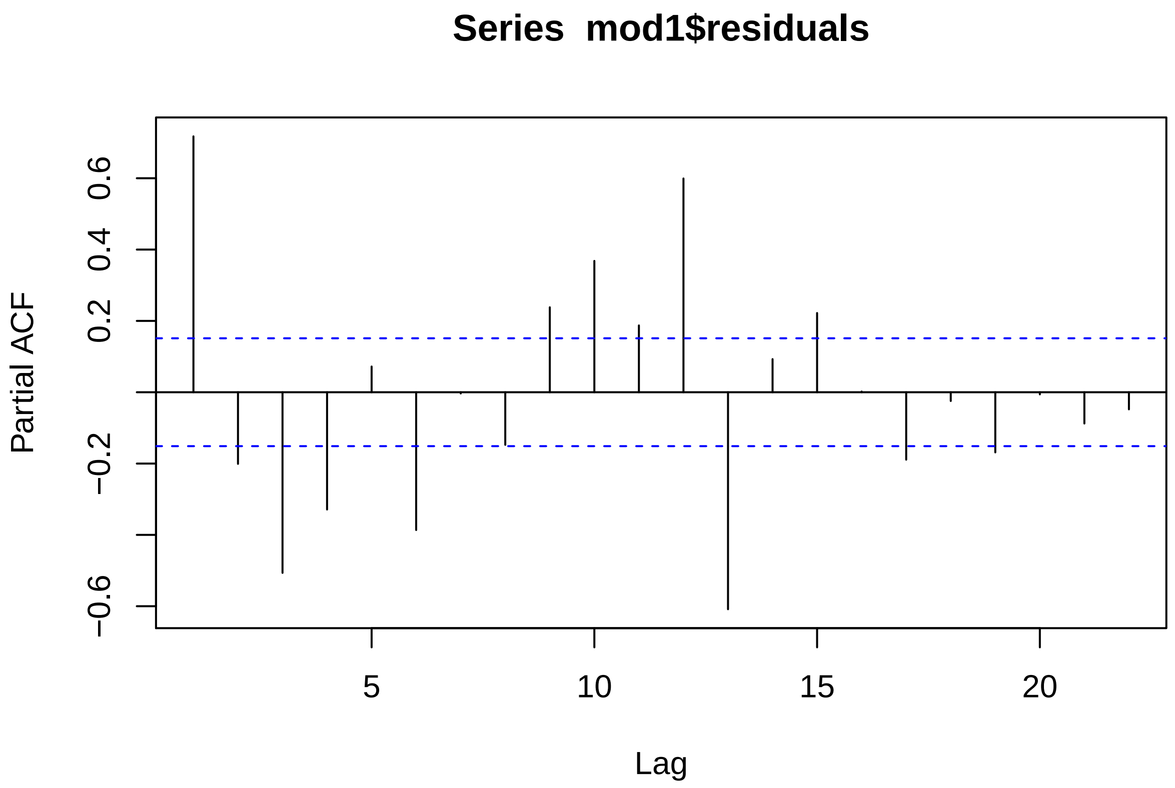
```
##
##  Jarque Bera Test
##
## data:  mod3$residuals
## X-squared = 6.3741, df = 2, p-value = 0.04129
##
##
##  Skewness
##
## data:  mod3$residuals
## statistic = 0.45292, p-value = 0.01655
##
##
##  Kurtosis
##
## data:  mod3$residuals
## statistic = 3.3001, p-value = 0.4272
```

1h

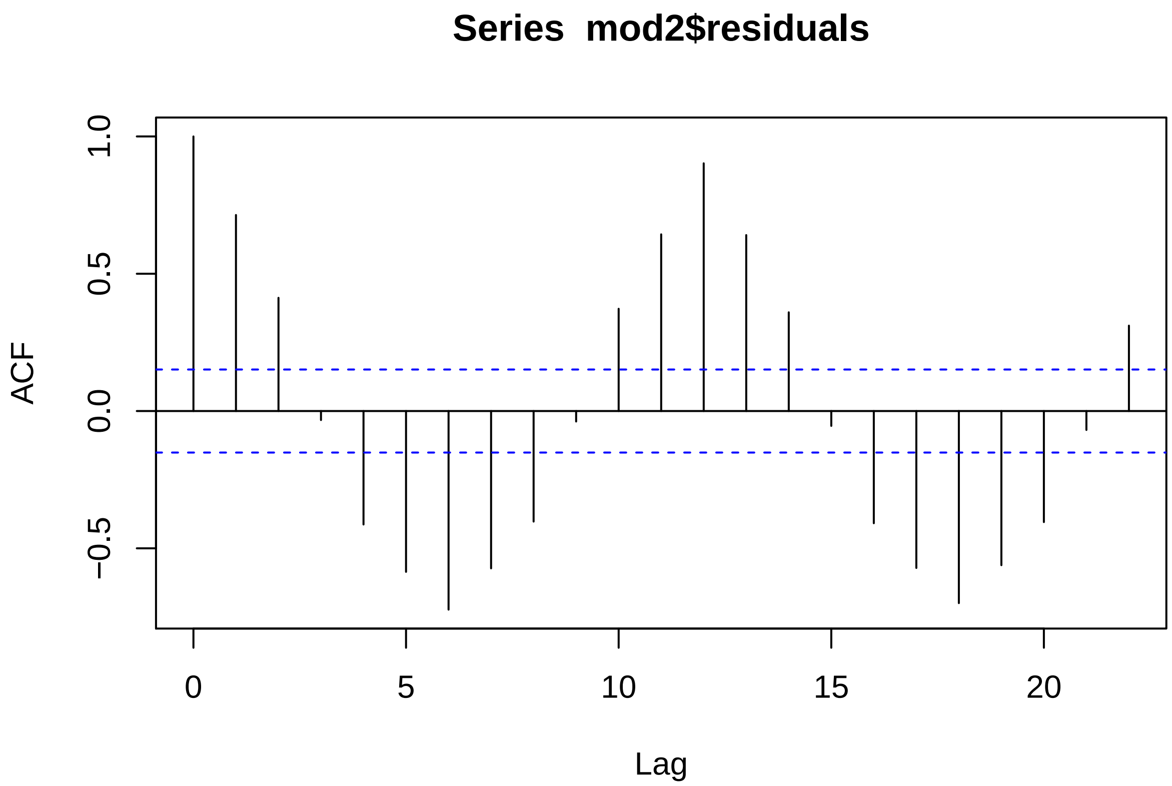
```
acf(mod1$residuals)
```



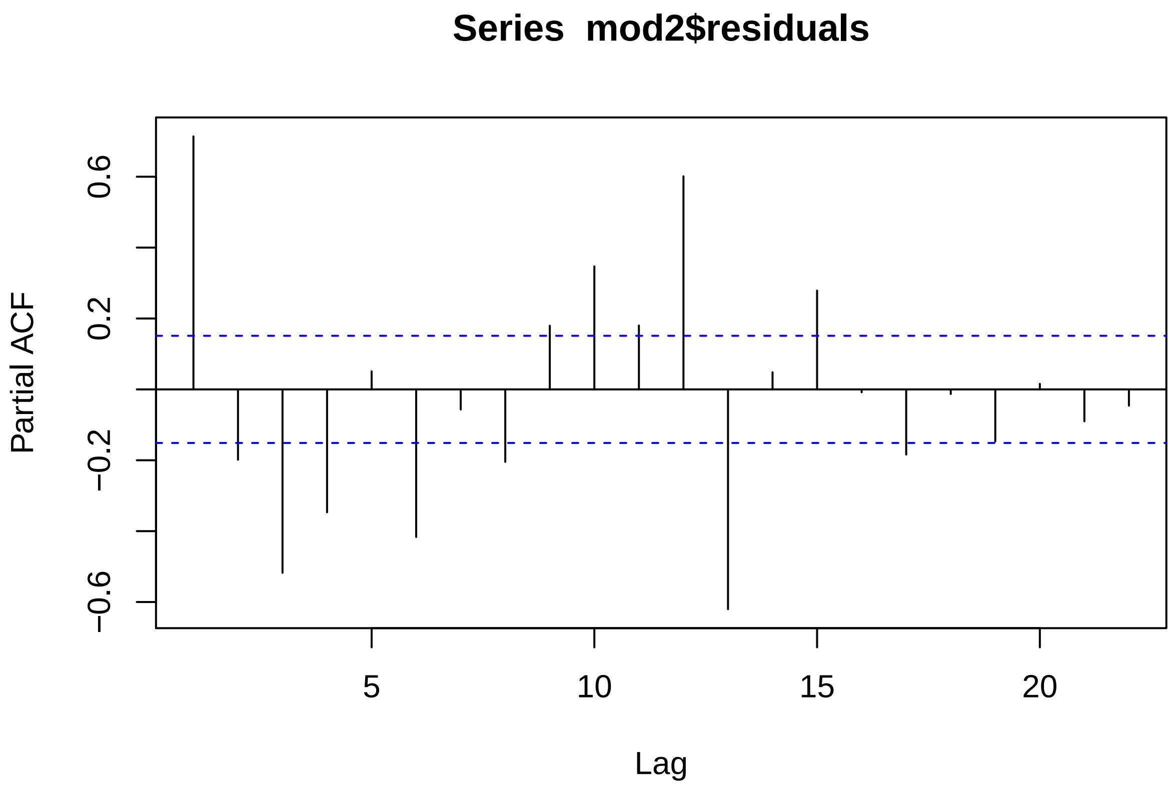
```
pacf(mod1$residuals)
```



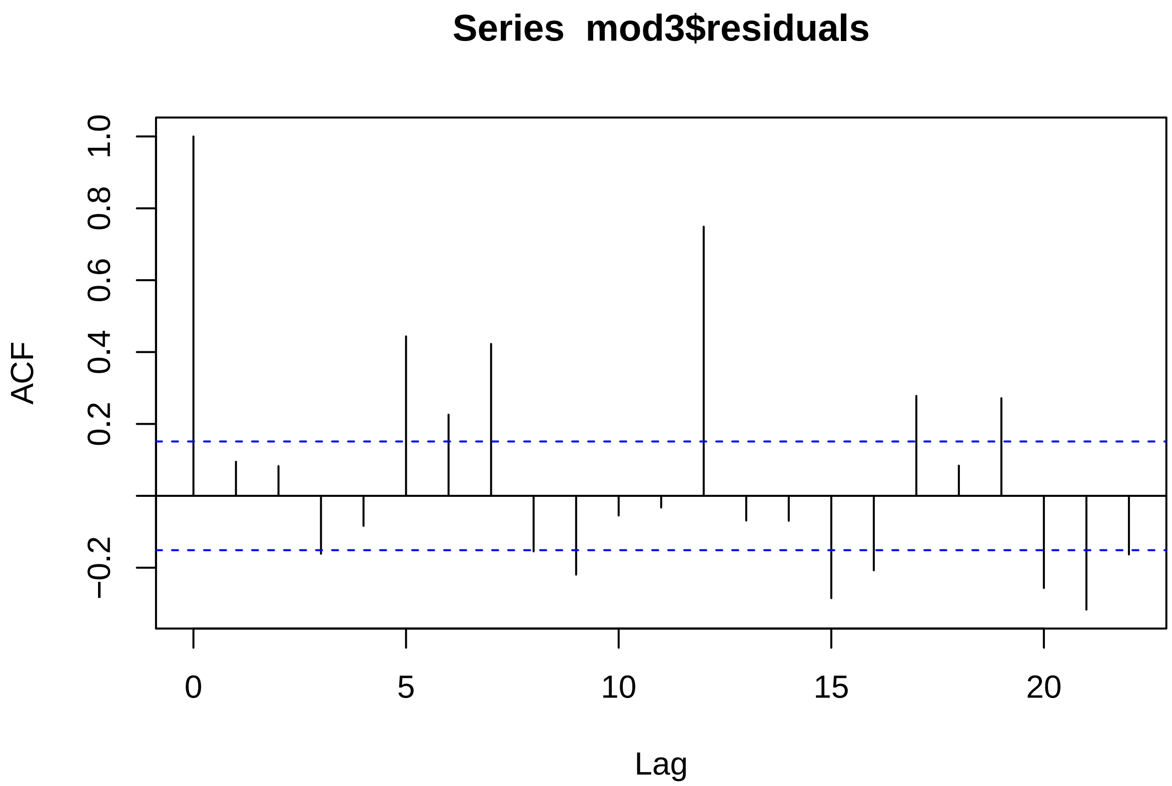
```
acf(mod2$residuals)
```



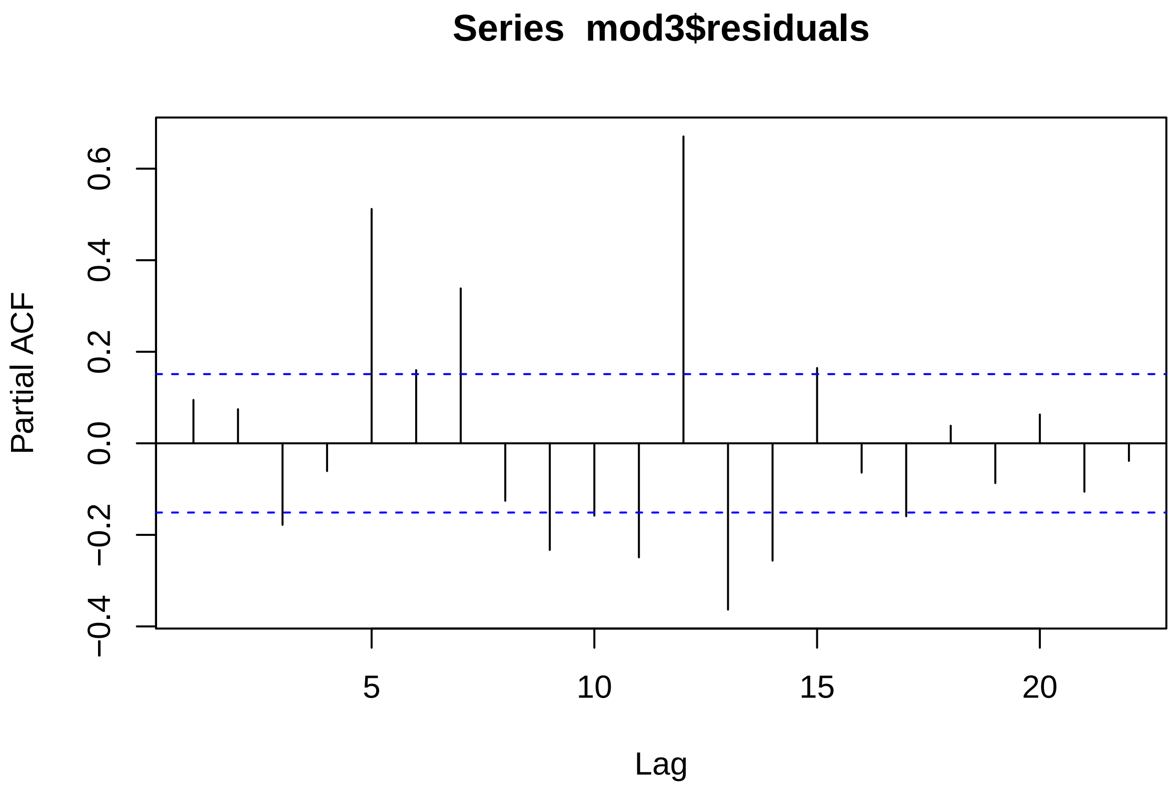
```
pacf(mod2$residuals)
```



```
acf(mod3$residuals)
```



```
pacf(mod3$residuals)
```



1i

(R2, t???distribution, F???distribution,etc.) TODO

```
summary(mod1)
```

```
##
## Call:
## lm(formula = datats ~ time)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -101.04  -50.02  -15.30   42.88  139.05
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -39237.56    2283.13  -17.19   <2e-16 ***
## time            20.31       1.16   17.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.74 on 166 degrees of freedom
## Multiple R-squared:  0.6489, Adjusted R-squared:  0.6468
## F-statistic: 306.8 on 1 and 166 DF,  p-value: < 2.2e-16
```

```
summary(mod2)
```

```
##
## Call:
## lm(formula = datats ~ time + I(time^2))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -92.95 -51.66 -15.73  44.30 134.94
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.234e+06  1.236e+06  -1.808   0.0725 .
## time         2.250e+03  1.255e+03   1.792   0.0749 .
## I(time^2)   -5.661e-01  3.187e-01  -1.776   0.0776 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.35 on 165 degrees of freedom
## Multiple R-squared:  0.6555, Adjusted R-squared:  0.6513
## F-statistic:   157 on 2 and 165 DF,  p-value: < 2.2e-16
```

```r
summary(mod3)
```

```
##
## Call:
## lm(formula = datats ~ time + I(time^2) + sin.t + cos.t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -83.542 -17.083   3.409  18.733  57.773
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.183e+06  5.553e+05  -3.932 0.000124 ***
## time         2.198e+03  5.640e+02   3.896 0.000142 ***
## I(time^2)   -5.528e-01  1.432e-01  -3.859 0.000164 ***
## sin.t        6.301e+01  2.963e+00  21.263  < 2e-16 ***
## cos.t       -4.201e+01  2.959e+00 -14.198  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.12 on 163 degrees of freedom
## Multiple R-squared:  0.9313, Adjusted R-squared:  0.9296
## F-statistic: 552.3 on 4 and 163 DF,  p-value: < 2.2e-16
```

1j

TODO

```r
AIC(mod1,mod2,mod3)
```

```
##      df      AIC
## mod1  3 1860.572
## mod2  4 1859.390
## mod3  6 1592.557
```

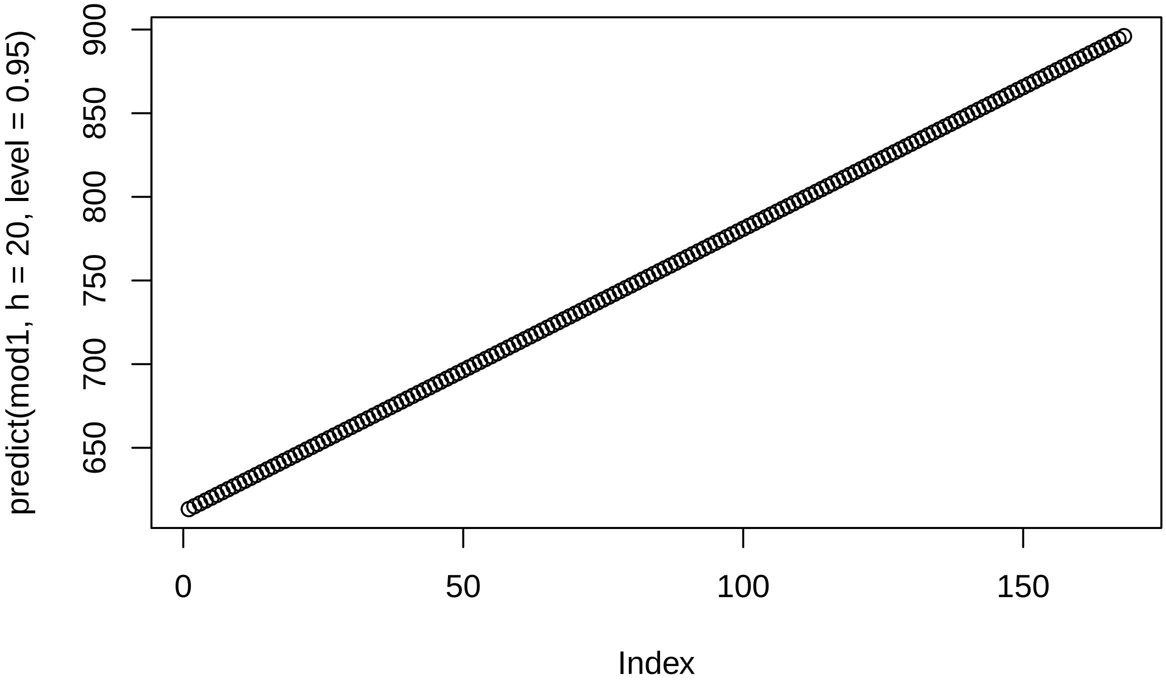
```
BIC(mod1,mod2,mod3)
```

```
##      df      BIC
## mod1  3 1869.944
## mod2  4 1871.886
## mod3  6 1611.301
```

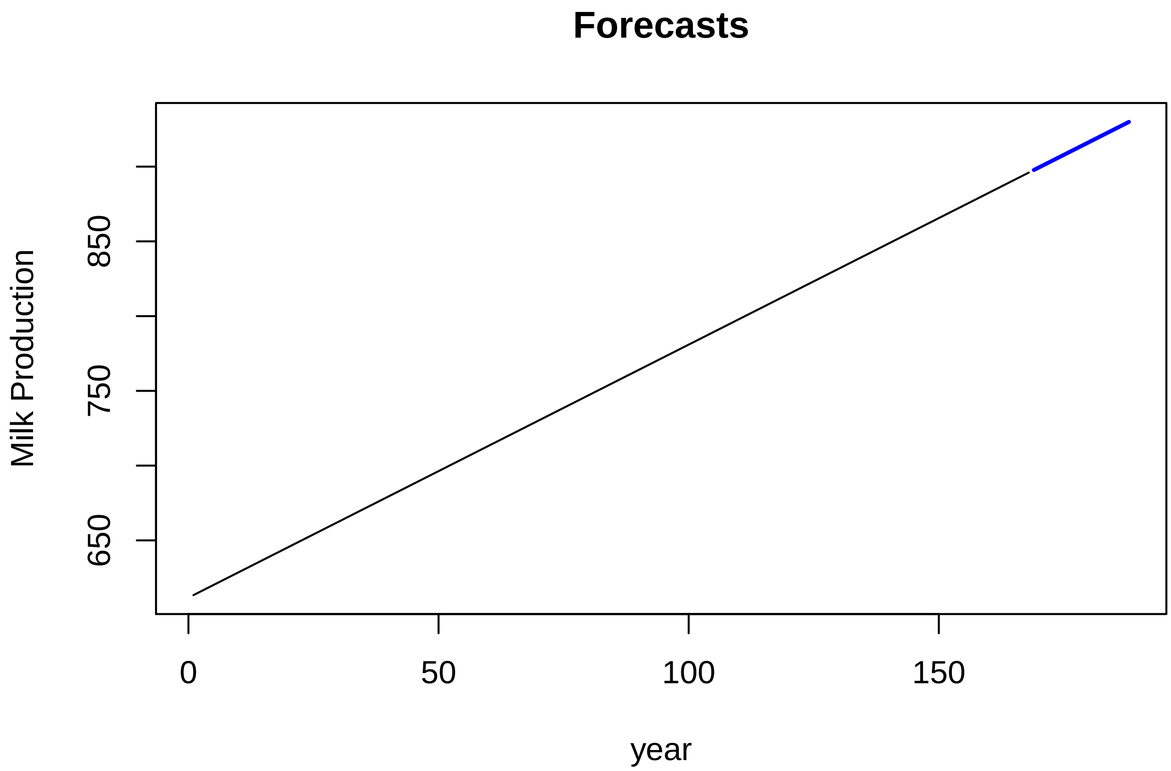
1k

TODO do all

```
library(forecast)
plot(predict(mod1,h = 20,level=.95))
```



```
plot(forecast(object = mod1$fitted.values,h = 20),main="Forecasts",xlab="year", ylab="Milk Production",s
```



2a

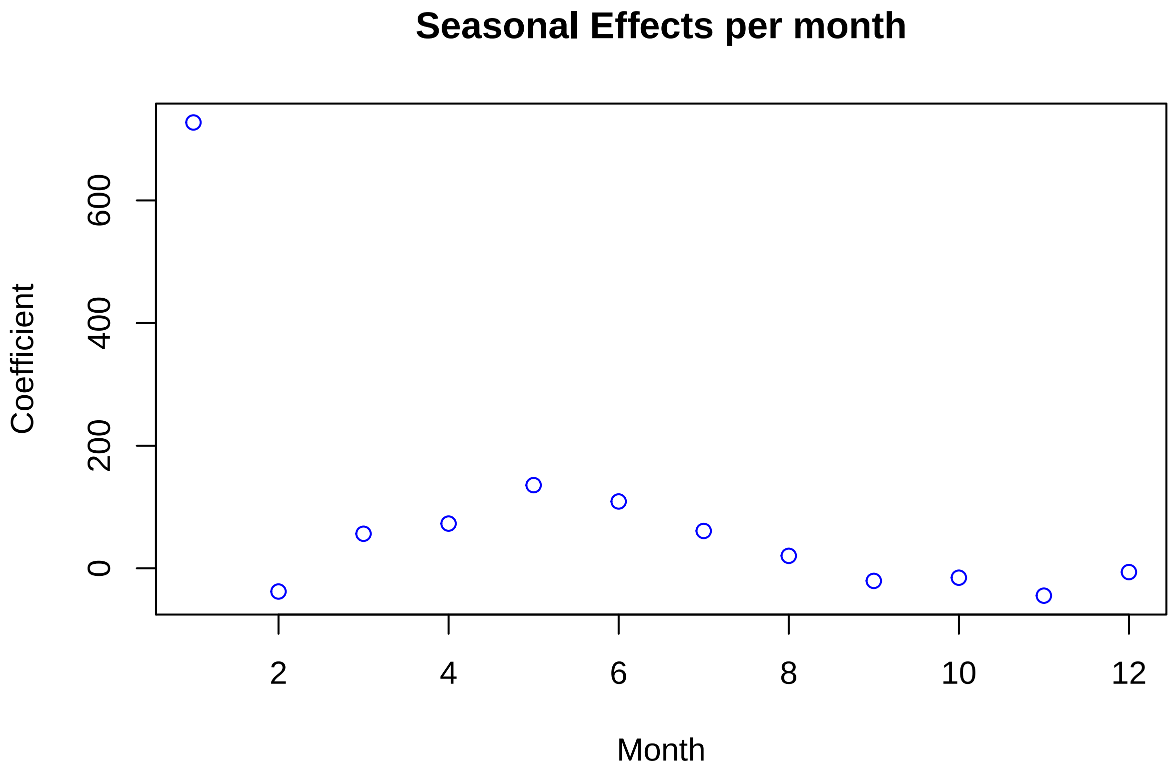
```
#seasonal dummies
seasonal=tslm(datats~season)
summary(seasonal)
```

```
##
## Call:
## tslm(formula = datats ~ season)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -148.500  -80.250    1.107   86.839  122.000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  727.071     23.576  30.840  < 2e-16 ***
## season2      -37.786     33.341  -1.133  0.25883
## season3       56.429     33.341   1.692  0.09256 .
## season4       73.000     33.341   2.189  0.03005 *
## season5      135.714     33.341   4.070 7.44e-05 ***
## season6      109.071     33.341   3.271  0.00132 **
## season7       61.000     33.341   1.830  0.06923 .
## season8       20.429     33.341   0.613  0.54096
## season9      -20.429     33.341  -0.613  0.54096
## season10     -15.214     33.341  -0.456  0.64880
## season11     -44.500     33.341  -1.335  0.18393
## season12      -6.071     33.341  -0.182  0.85574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.21 on 156 degrees of freedom
## Multiple R-squared:  0.3041, Adjusted R-squared:  0.2551
## F-statistic: 6.198 on 11 and 156 DF,  p-value: 2.055e-08
```

2b

TODO INTERPERET

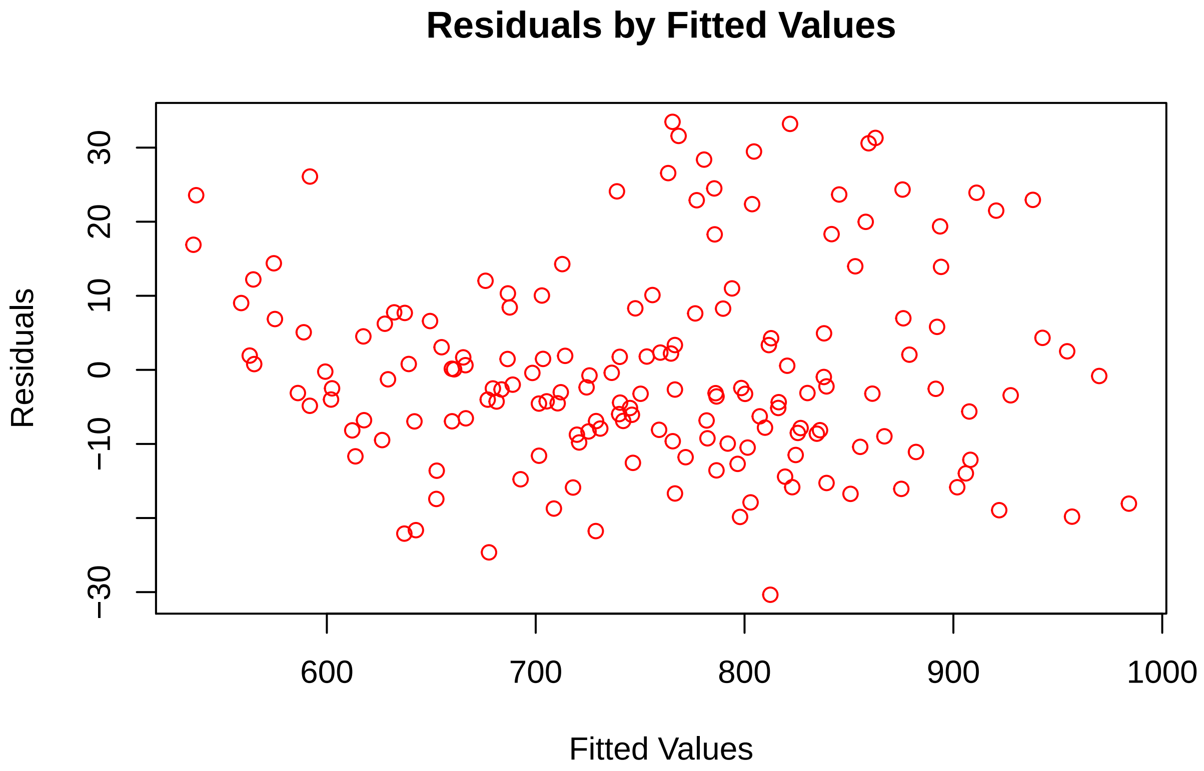
```r
#plot factors (I dont think this is what he wants)
plot(seasonal$coefficients,col='blue',xlab='Month',ylab='Coefficient',main='Seasonal Effects per month')
```



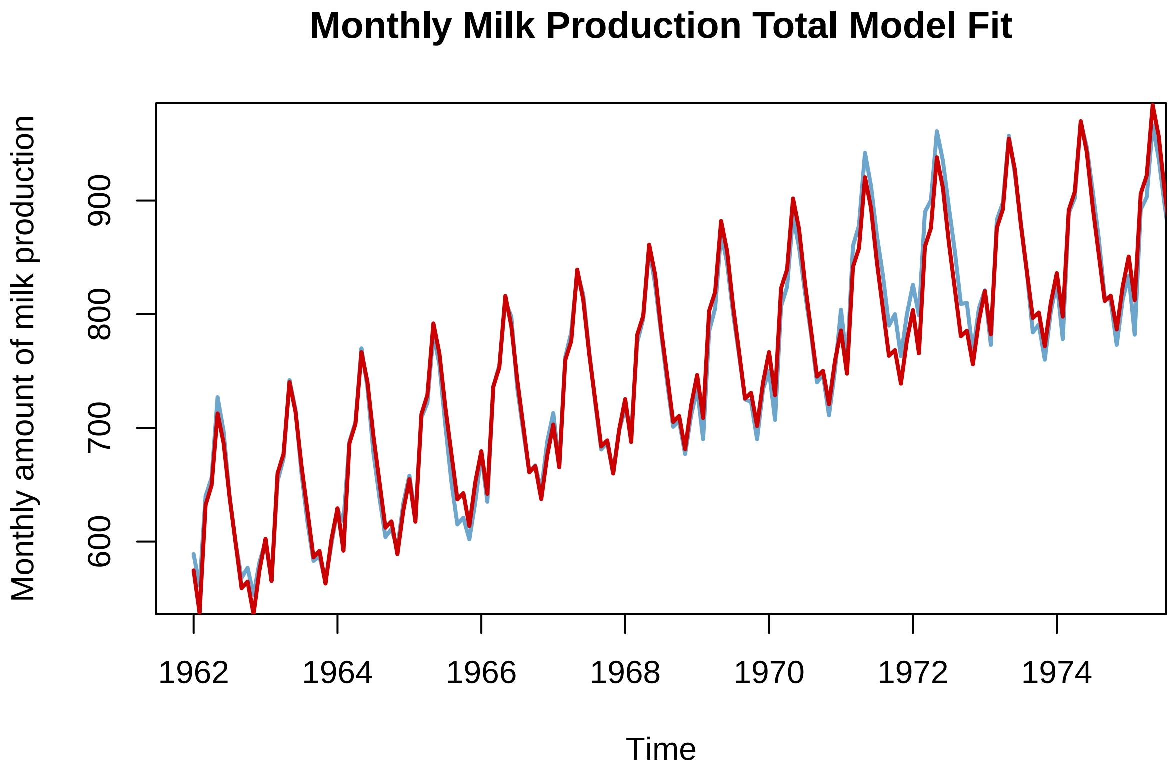
2c

TODO DISCUSS

```r
seasontrend = tslm(datats~time + I(time^2) + sin.t + cos.t + season)
plot(seasontrend$fitted.values,seasontrend$residuals,col='red',xlab='Fitted Values',ylab='Residuals',ma
```



```
plot(datats,ylab="Monthly amount of milk production", xlab="Time", lwd=2, col='skyblue3', xlim=c(1962,19
lines(time,seasontrend$fitted.values,col="red3",lwd=2)
```



2d

TODO DISCUSS

```
summary(seasontrend)
```

```
##
## Call:
## tslm(formula = datats ~ time + I(time^2) + sin.t + cos.t + season)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.354  -8.537  -2.532   6.888  33.480
##
## Coefficients: (2 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.187e+06  2.829e+05  -7.730 1.30e-12 ***
## time         2.201e+03  2.874e+02   7.660 1.93e-12 ***
## I(time^2)   -5.537e-01  7.298e-02  -7.588 2.89e-12 ***
## sin.t        3.177e+01  1.749e+01   1.817  0.07119 .
## cos.t        6.860e+01  2.936e+01   2.336  0.02075 *
## season2     -4.625e+01  1.520e+01  -3.042  0.00276 **
## season3      5.969e+01  3.149e+01   1.896  0.05989 .
## season4      1.046e+02  4.805e+01   2.176  0.03108 *
## season5      2.041e+02  6.021e+01   3.389  0.00089 ***
## season6      2.124e+02  6.467e+01   3.285  0.00126 **
## season7      1.877e+02  6.021e+01   3.118  0.00217 **
## season8      1.521e+02  4.804e+01   3.167  0.00186 **
## season9      9.607e+01  3.148e+01   3.052  0.00268 **
## season10     6.954e+01  1.520e+01   4.576 9.68e-06 ***
## season11            NA         NA      NA       NA
## season12            NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.82 on 154 degrees of freedom
## Multiple R-squared:  0.9831, Adjusted R-squared:  0.9817
## F-statistic:   691 on 13 and 154 DF,  p-value: < 2.2e-16
```

2e

TODO DISCUSS

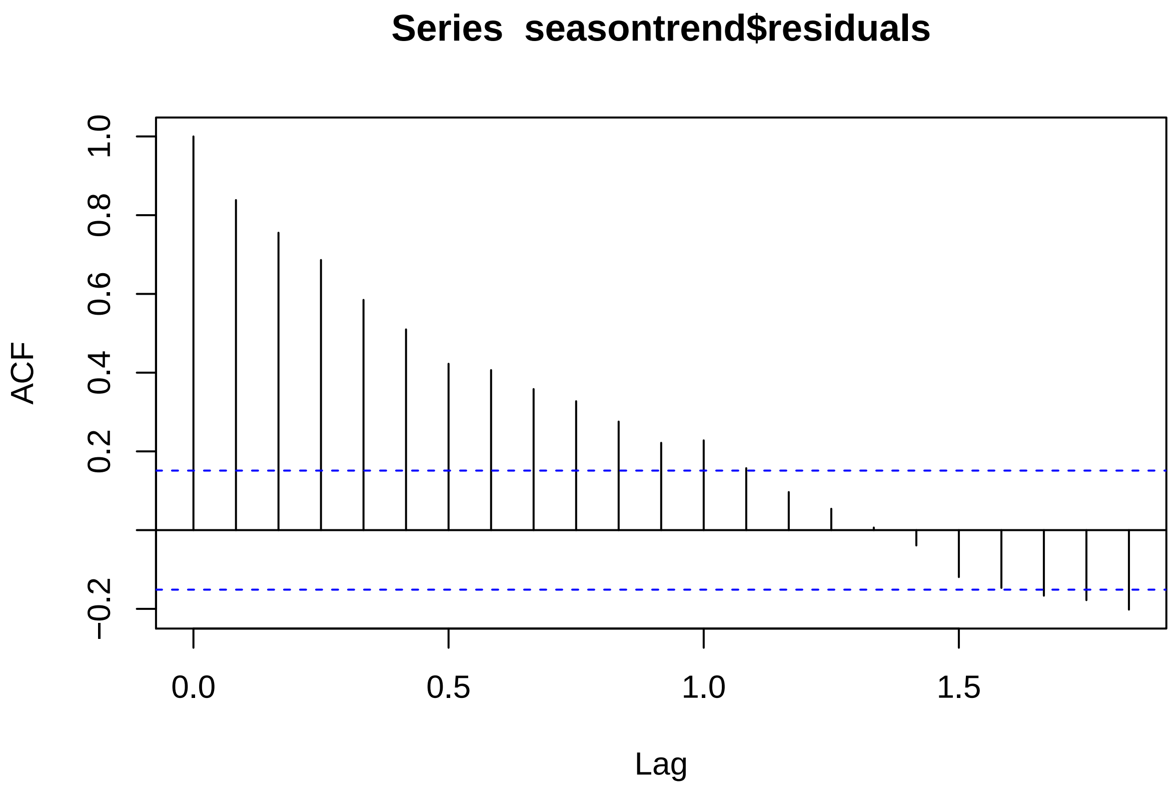
```
JarqueBera.test(seasontrend$residuals)
```

```
##
##  Jarque Bera Test
##
## data:  seasontrend$residuals
## X-squared = 10.834, df = 2, p-value = 0.00444
##
##
##  Skewness
##
## data:  seasontrend$residuals
## statistic = 0.62193, p-value = 0.0009985
##
##
##  Kurtosis
##
## data:  seasontrend$residuals
## statistic = 2.9763, p-value = 0.9499
```

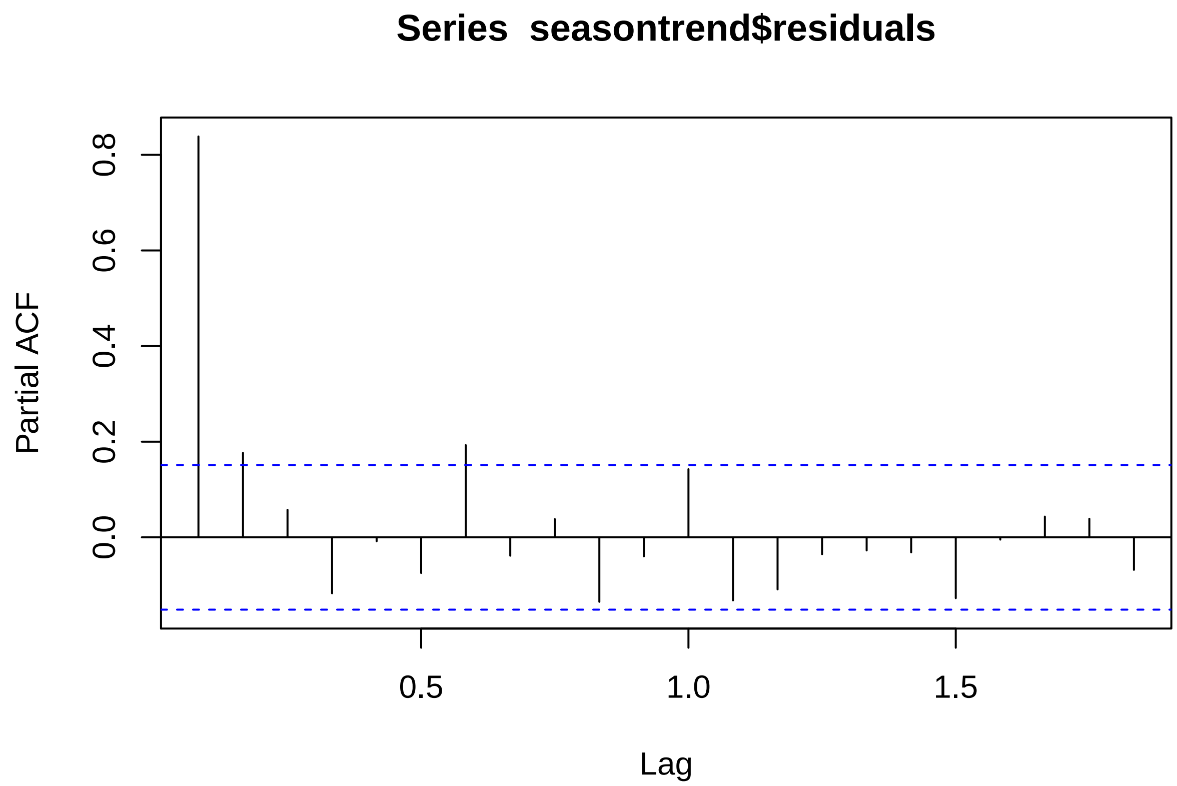
2f

TODO DISCUSS

```
acf(seasontrend$residuals)
```



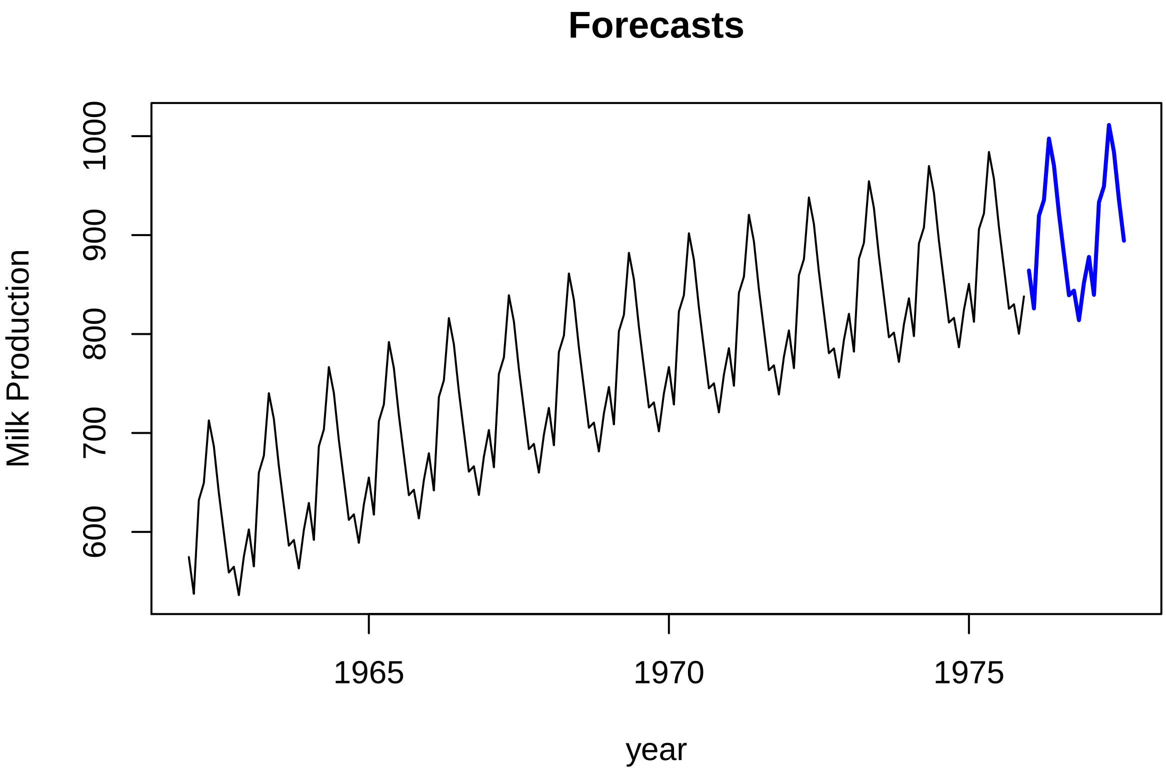
```
pacf(seasontrend$residuals)
```



2g

TODO PREDICTION INTERVAL

```
plot(forecast(object = seasontrend$fitted.values,h = 20),main="Forecasts",xlab="year", ylab="Milk Produ
```



III

Conclusions and Future Works

Our final model included a dummy variable for each month, a periodic term, and a quadtraic term over time. With these simple features, we are able to have a fitted model that matches the data almost exactly. We've considered some possible work we can add to this:

It would be interested to get data after 1975 and use our model to make predictions on it. We can also use this to make future predctions in 2019. Additionally, our model doesn't fit the data perfectly. There may be some other predictors we can use when we fit our model.

IV

References

https://datamarket.com/data/set/22ox/monthly-milk-production-pounds-per-cow-jan-62-dec-75#!ds=22ox&display=line

Agriculture, Source: Cryer (1986), in file: data/milk, Description: Monthly milk production: pounds per cow. Jan 62 - Dec 75

V

R Code

included