Week 4 Homework

Matthew Dunne

February 1, 2019

Week 4: Homework Assignment

This assignment helps understanding stationarity and seasonality of linear models for time series

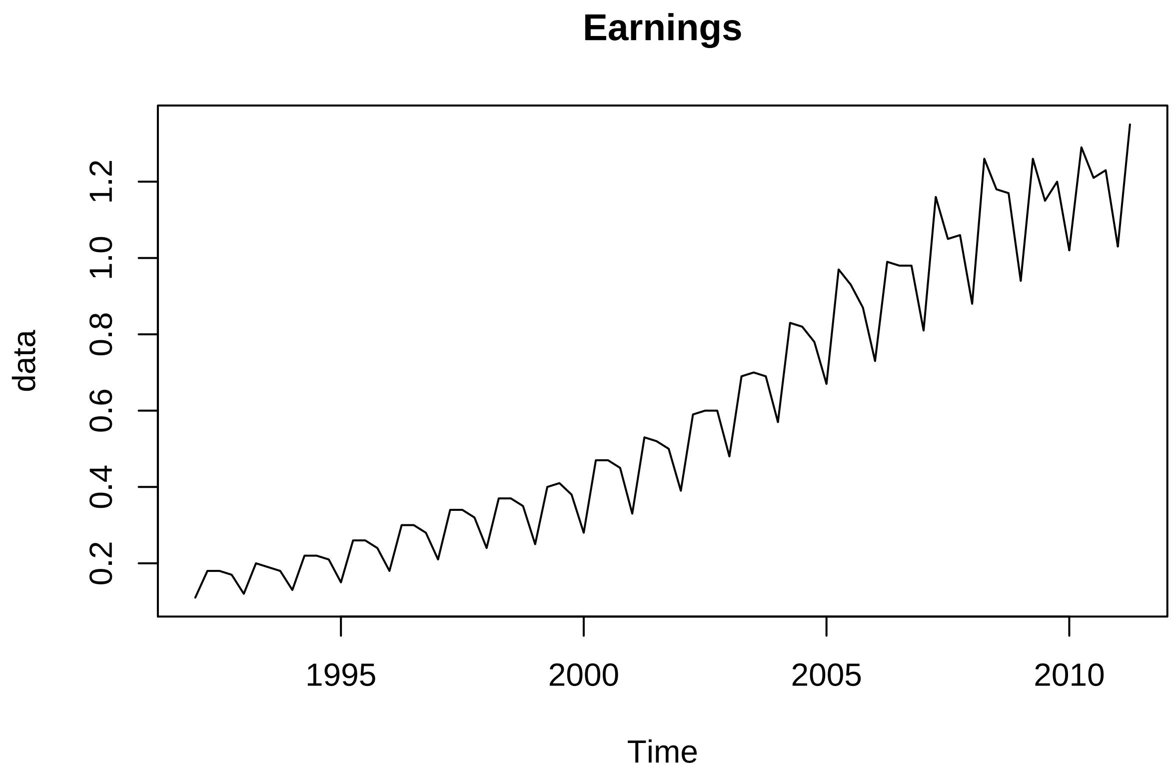
Exercise 7 on page 126 of the Textbook

Consider the quarterly earnings per share of Johnson & Johnson from the first quarter of 1992 to the second quarter of 2011. The data are in the file q-jnj-earns-9211.txt available on the textbook web page. Take log transformation if necessary.

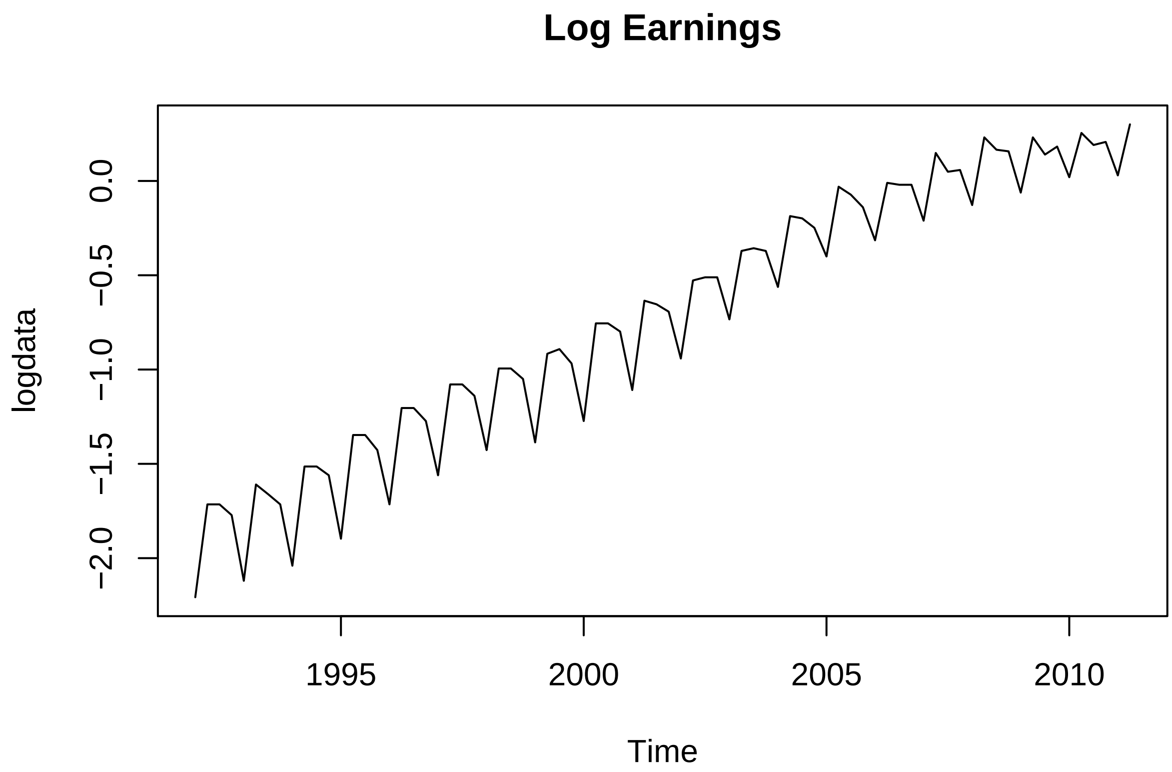
```
suppressWarnings(library(tseries))
suppressMessages(library(fpp))
suppressMessages(library(TSA))
datapath<-"C:/Users/mjdun/Desktop/Financial Analytics/Week 4"
data=read.table(file=paste(datapath,"q-jnj-earns-9211.txt",sep="/"),header=T)
#take out the relevant column
data_actual=data$earns
#convert to time series
data=ts(data_actual,frequency=4,start=c(1992,1))
logdata=log(data)
```

Should we take a log transformation? Let us see what the plots look like.

```
plot(data, main='Earnings')
```




```
plot(logdata, main='Log Earnings')
```



Without taking the log, earnings look exponential rather than linear and have different levels of volatility at different parts of the time series. With the log transformation there is still some difference in volatility, with the later data now being less volatile than the earlier data. *So use the log transformation.*

Build a time series model for the data.

There is clear seasonality (quarterly) and a linear trend. We will need to use differencing to account for both.

Let us try two different models: the first using the auto.arima function, the second using the airline model.

```
#let auto.arima find something
model1<-auto.arima(logdata, seasonal=TRUE)
#use the airline model. page 101 of book, page 12 of workshop
model2<-arima(logdata,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
```

Perform model checking to assess the adequacy of the fitted model.

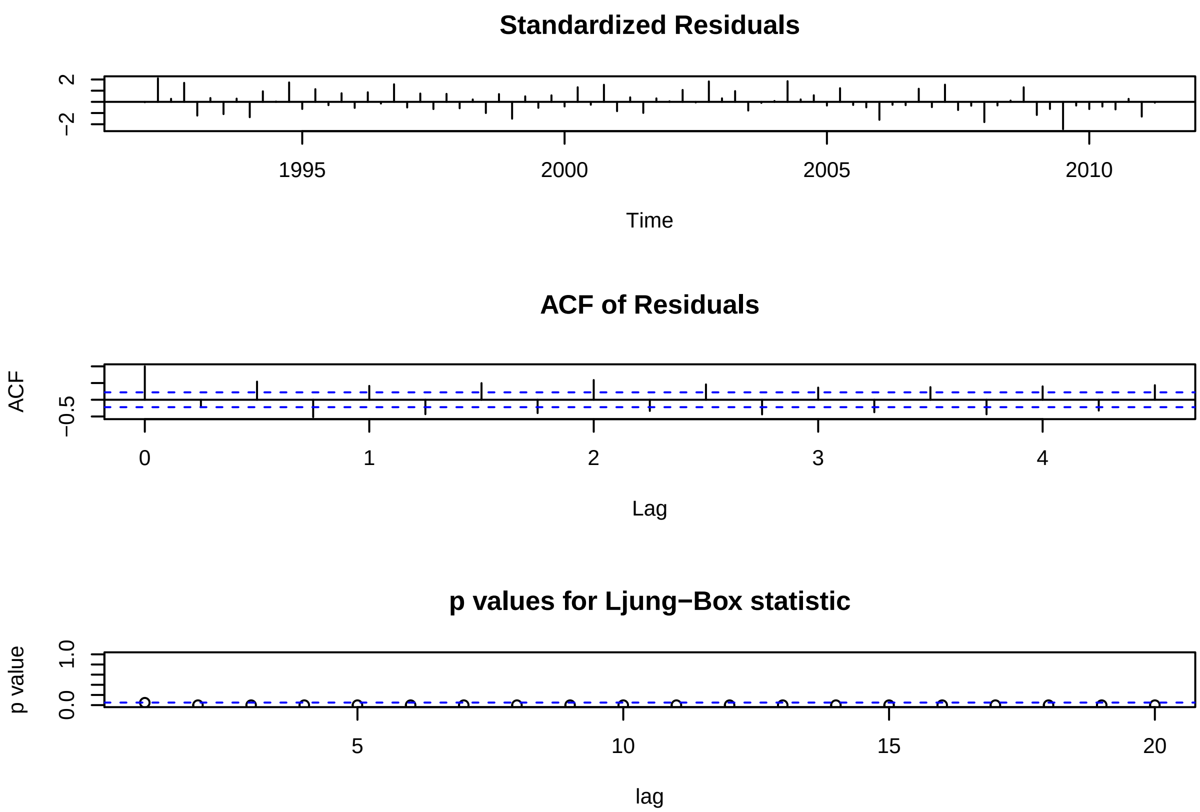
First check the auto.arima model

```
model1
```

```
## Series: logdata
## ARIMA(2,1,3)(0,0,1)[4] with drift
##
## Coefficients:
##          ar1      ar2      ma1     ma2      ma3    sma1   drift
##       0.0100  -0.9724  -1.0372  1.1893  -0.7256  0.8023  0.0278
## s.e.  0.0262   0.0243   0.0932  0.1177   0.0960  0.0643  0.0030
##
## sigma^2 estimated as 0.004709:  log likelihood=95.02
```

```
## AIC=-174.05   AICc=-171.93   BIC=-155.3
```

```r
tsdiag(model1,gof=20)
```

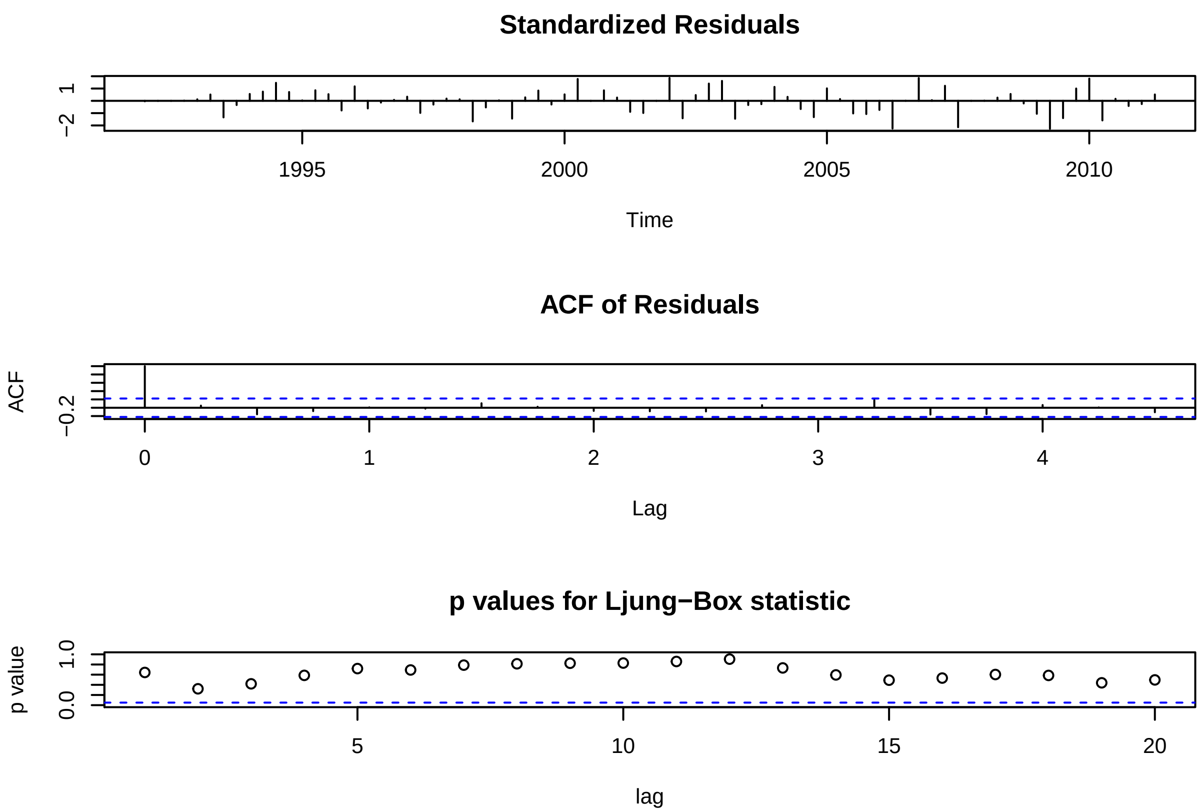


This model does not perform well in terms of what we would expect to see in the residuals.

- top plot: The standardized residuals have many large and many small values. We would expect them to be i.i.d.
- middle plot: There is still autocorrelation in the residuals.
- bottom plot: Pretty much all the p-values are below 0.05, meaning we reject the null hypothesis that the residuals are white noise.

Now check the airline model.

```r
tsdiag(model2,gof=20)
```



- top plot: The standardized residuals look more or less i.i.d.
- middle plot: There is no autocorrelation in the residuals (no spikes after 0)
- bottom plot: All the p-values are above 0.05, meaning we *do not reject* the null hypothesis that the residuals are white noise.

So we will use the airline model.

Write down the model.

```
model2
```

```
##
## Call:
## arima(x = logdata, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 4))
##
## Coefficients:
##           ma1     sma1
##       -0.3223  -0.2175
## s.e.   0.1372   0.1208
##
## sigma^2 estimated as 0.0011:  log likelihood = 144.9,  aic = -285.81
```

Written differently, the estimated model is:

$$(1-B)(1-B^4)\xi_t = (1-0.3223B)(1-0.2175B^4)\epsilon_t$$

and

$$\sigma_\epsilon^2 = 0.0011$$

We can also look at the Box Test.

```
Box.test(model2$residuals, lag=12, type = 'Ljung')
```

```
##
##  Box-Ljung test
##
## data:  model2$residuals
## X-squared = 6.2041, df = 12, p-value = 0.9054
```

We would not reject the null hypothesis of white noise in the residuals. This is true even when we take into account degrees of freedom.

```
pp=1-pchisq(6.2041,10)
pp
```

```
## [1] 0.797834
```

Refit the model using data from 1992 to 2008.

```
refit_data<-window(logdata, c(1992, 1), c(2008, 4))
refit_model<-arima(refit_data,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
```

Perform 1-step to 10-step forecasts of quarterly earnings and obtain a forecast plot.

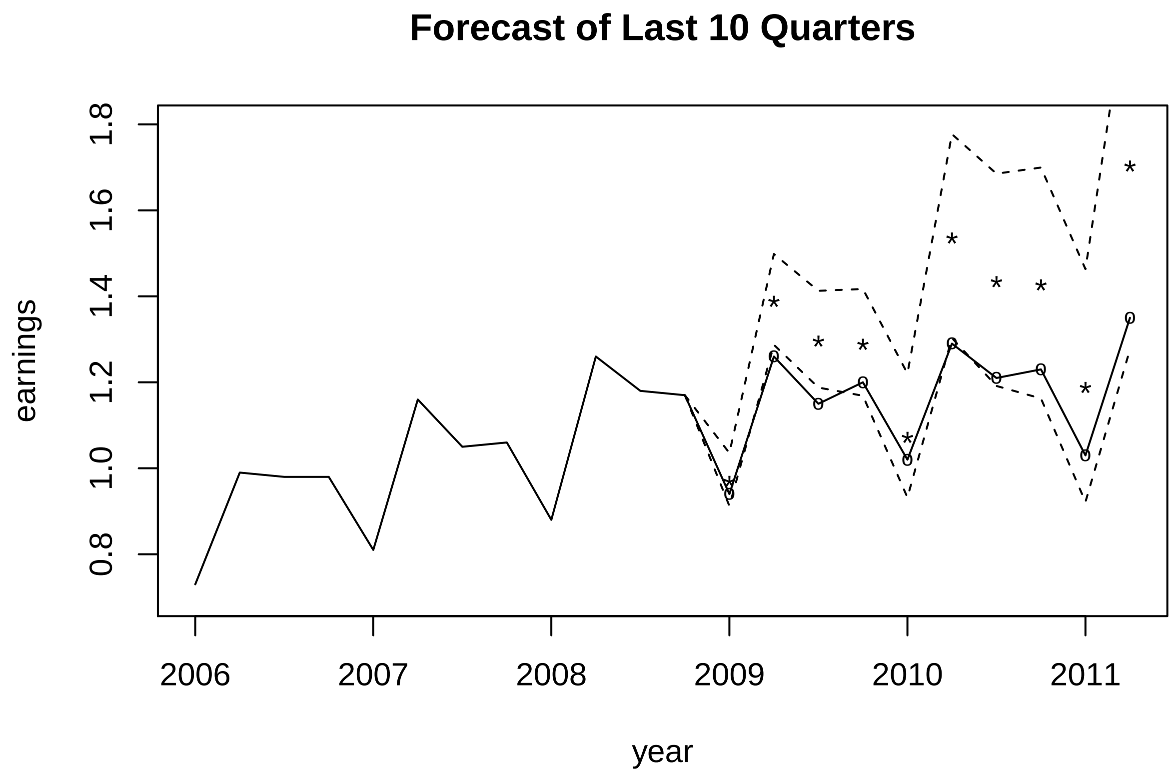
```
prediction<-predict(refit_model,10)
#predictions into a vector
pred<-prediction$pred
#standard errors into a vector
se<-prediction$se
#point forecasts de-logged
forecasts<-exp(pred+se^2/2)
#get standard deviation of the forecast error
v1<-exp(2*pred+se^2)*(exp(se^2)-1)
s1<-sqrt(v1)
#actual values to include in the chart, for the 10 you are predicting plus three years before that
#use the original, non-log data because we left out the last 10 for retraining the model
eps<-data_actual[57:78]
#how many timer periods are you going to chart
length(eps)
```

```
## [1] 22
```

```
#time index. start at beginning of 2006 (2005+1.0, 2005+1.25 ...)
tdx<-(c(1:22)+3)/4+2005
#upper and lower bounds, starting at last observation before we start forecasting
upp<-c(data_actual[68], forecasts+2*s1)
low<-c(data_actual[68], forecasts-2*s1)
```

Plot it.

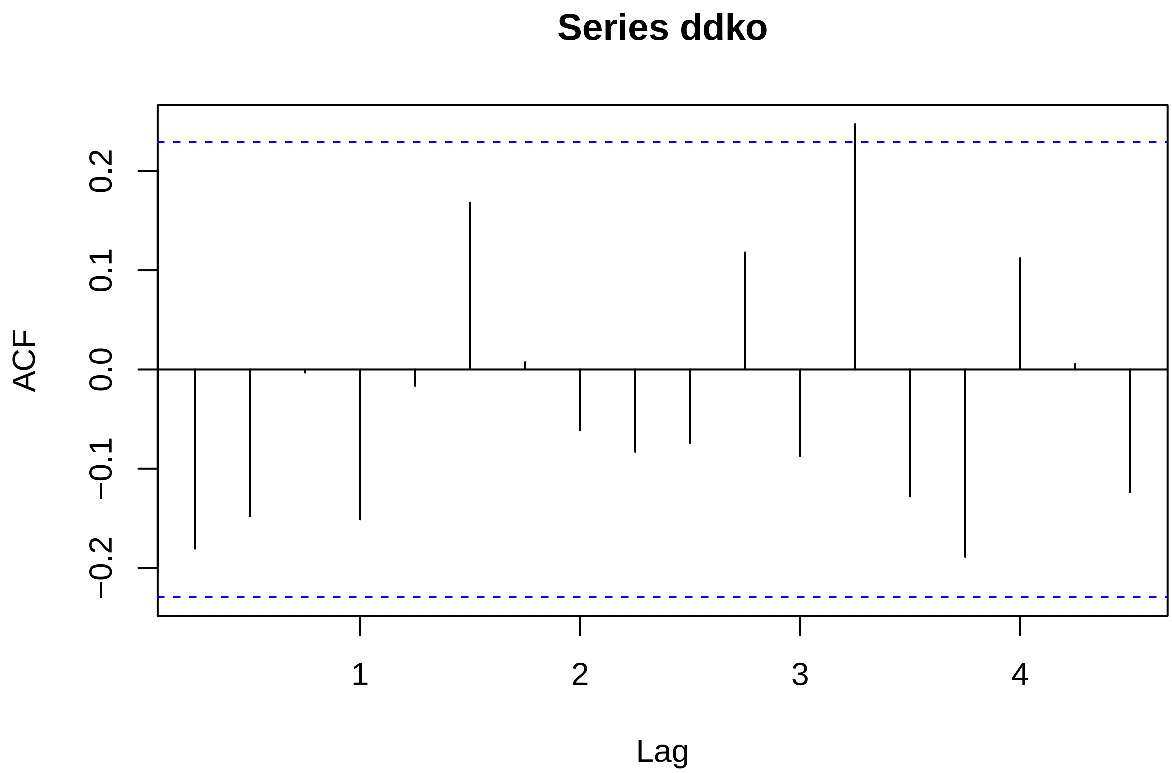
```
plot(tdx,eps,xlab='year',ylab='earnings',type='l', ylim=c(0.7, 1.8), main='Forecast of Last 10 Quarters
points(tdx[13:22],forecasts,pch='*')
lines(tdx[12:22],upp,lty=2)
lines(tdx[12:22],low,lty=2)
points(tdx[13:22], data_actual[69:78],pch='o',cex=0.7)
```



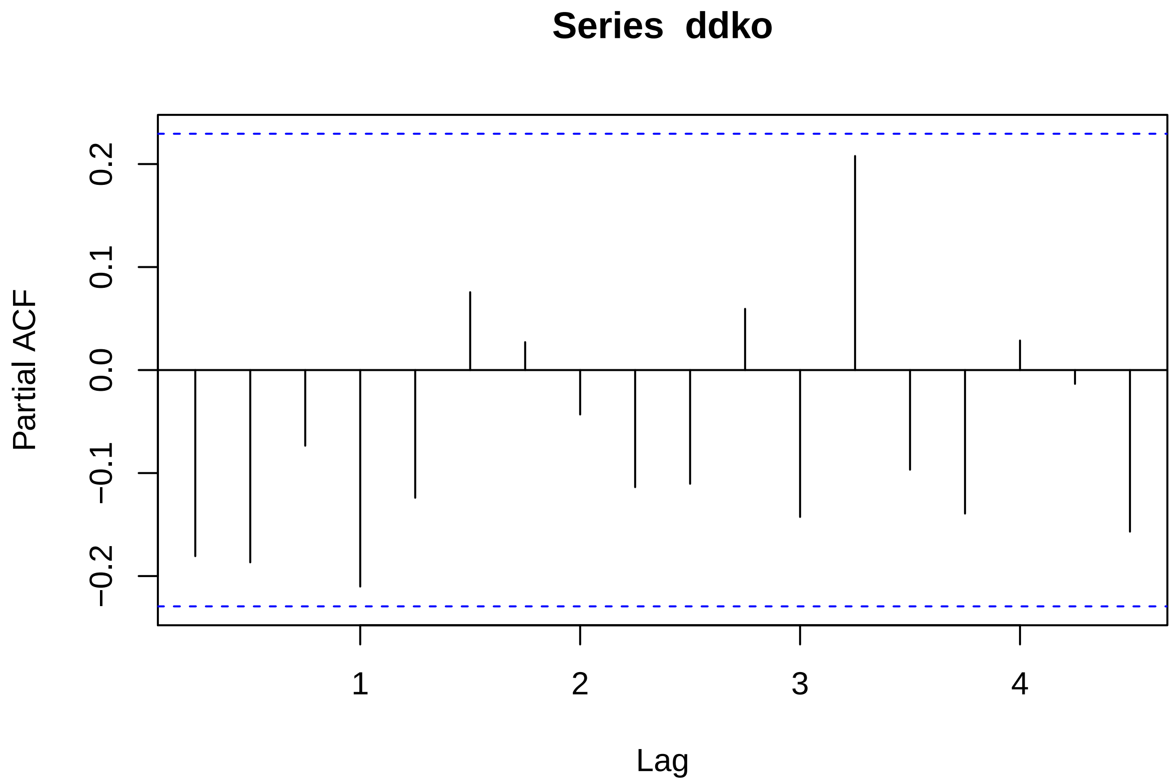
We see that this model over-estimates earnings for the forecast period. This is not surprising when we look at the Log Earnings chart and see that the earnings become less volatile at pretty much the point we start forecasting. The model is trained on data that does not capture this trend. This trend likely arises from the financial crisis in 2008, as this almost certainly affected corporate earnings.

Can we use intervention analysis to change the model? Let us take out the linear trend and the seasonality with differencing to see whether we can fit a SARIMA model on what is left.

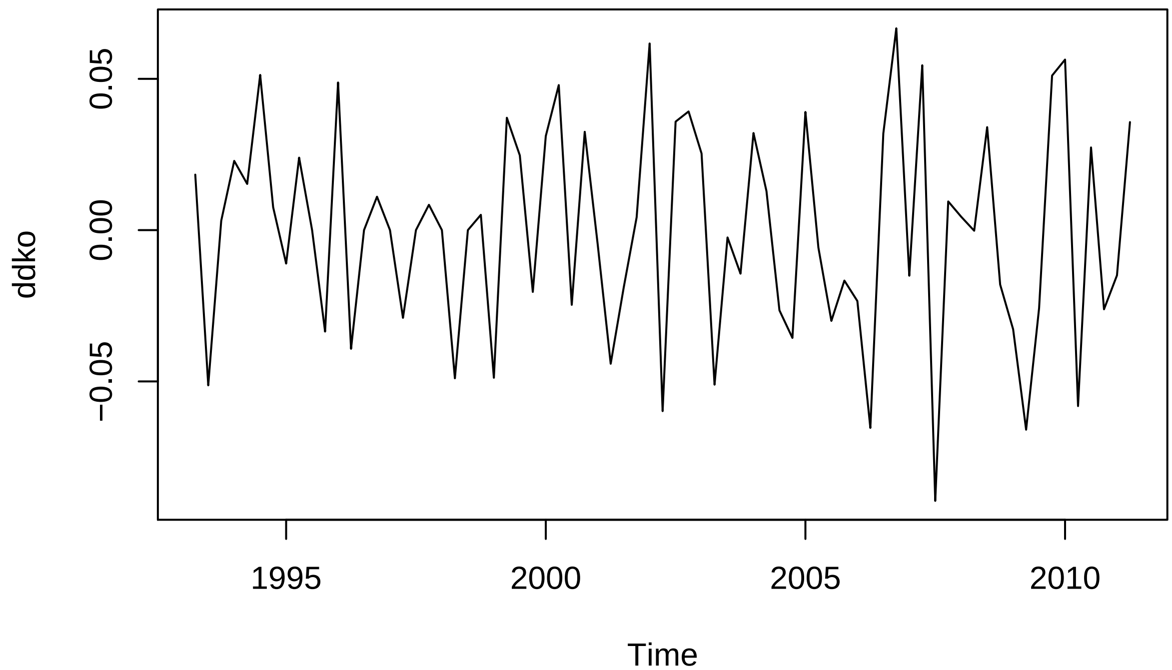
```
#take out linear trend
dko<-diff(logdata)
#take out seasonality
ddko<-diff(dko, 4)
acf(ddko)
```




```r
pacf(ddko)
```



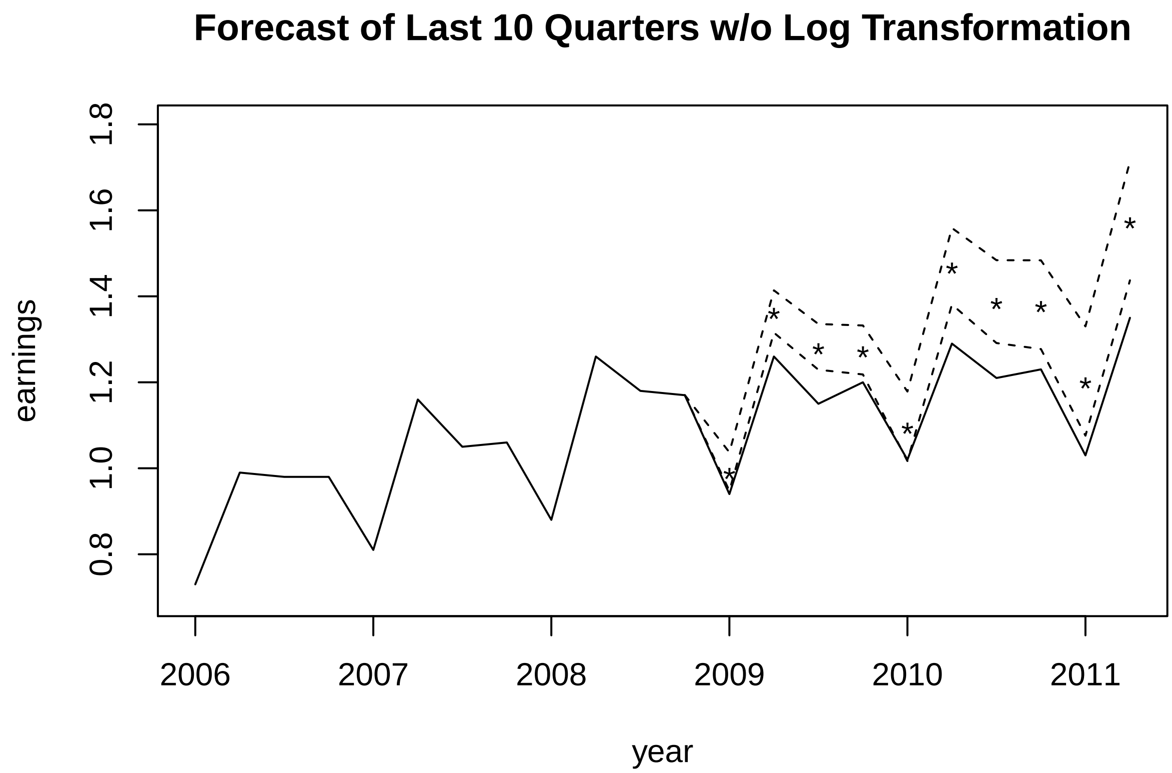

```r
plot(ddko)
```



What is left looks like white noise. No SARIMA model suggests itself for purposes of Intervention model estimation.

Interestingly if we use the original data (without log transformation) we see much the same result.

```
refit_data<-window(data, c(1992, 1), c(2008, 4))
refit_model<-arima(refit_data,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
prediction<-predict(refit_model,10)
pred<-prediction$pred
se<-prediction$se
upper<-c(data_actual[68], pred+1.96*se)
lower<-c(data_actual[68], pred-1.96*se)
plot(tdx,eps,xlab='year',ylab='earnings',type='l', ylim=c(0.7, 1.8), main='Forecast of Last 10 Quarters
points(tdx[13:22],pred,pch='*')
lines(tdx[12:22],upper,lty=2)
lines(tdx[12:22],lower,lty=2)
```


Forecast of Last 10 Quarters w/o Log Transformation
earnings
year
0.8
1.0
1.2
1.4
1.6
1.8
2006
2007
2008
2009
2010
2011