**Meenakshi Nagarajan**

**Project**

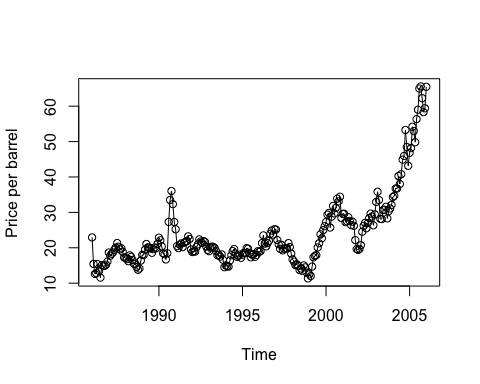
**STT4110/6110**

**1.** Our textbook uses monthly oil price series data set in many sections. Read those related sections. Compile all the R codes that appear in the textbook (from Chapter 1 to Chapter 9) for this data set into one complete R script file and analyze this data set following the thread of our textbook.

*Solution:*

CHAPTER 1

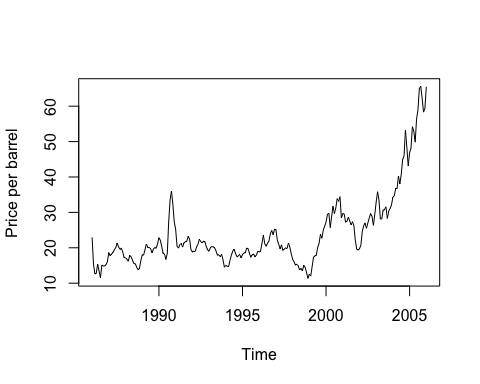
data("oil.price")  
plot(oil.price,type='o',ylab='Price per barrel')



*Comment:* *The time series plot of oil price data is shown here. Patterns in the data are not quite clear in this graph*

CHAPTER 2

plot(oil.price,type='l',ylab='Price per barrel')



*Comment: The same graph is displayed here from the year of January 1986 to January 2006. The series displays a variation from the year of 2000.*

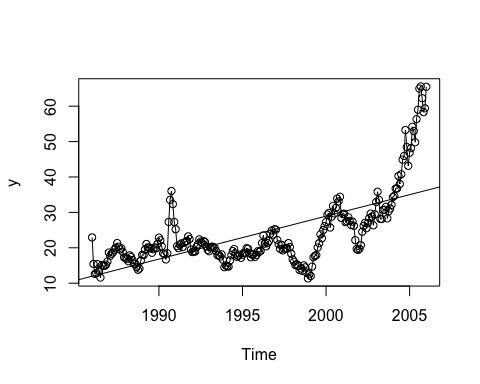
CHAPTER 3

model1=lm(oil.price~time(oil.price))  
summary(model1)

##   
## Call:  
## lm(formula = oil.price ~ time(oil.price))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.2726 -4.8340 -0.5501 3.0362 29.7837   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.396e+03 1.752e+02 -13.68 <2e-16 \*\*\*  
## time(oil.price) 1.212e+00 8.776e-02 13.82 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.898 on 239 degrees of freedom  
## Multiple R-squared: 0.444, Adjusted R-squared: 0.4417   
## F-statistic: 190.9 on 1 and 239 DF, p-value: < 2.2e-16

*Comment: The estimated slope and intercept are given in the model. Adjusted r square value is not too high which means that 44% variation in the oilprice is explained by the linear time trend.*

plot(oil.price,type='o',ylab='y')  
abline(model1)# add the fitted least squares line from model1



*Comment: This graph shows the time-series of oil price data with the least squares regression trend line superimposed.*

#Fitting seasonal means model to oilprice data  
month.=season(oil.price)  
model2=lm(oil.price~month.-1)  
summary(model2)

##   
## Call:  
## lm(formula = oil.price ~ month. - 1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.687 -6.213 -3.312 3.243 40.477   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## month.January 25.003 2.349 10.642 <2e-16 \*\*\*  
## month.February 22.534 2.407 9.360 <2e-16 \*\*\*  
## month.March 22.874 2.407 9.502 <2e-16 \*\*\*  
## month.April 22.983 2.407 9.547 <2e-16 \*\*\*  
## month.May 23.381 2.407 9.712 <2e-16 \*\*\*  
## month.June 23.362 2.407 9.704 <2e-16 \*\*\*  
## month.July 23.676 2.407 9.835 <2e-16 \*\*\*  
## month.August 24.966 2.407 10.371 <2e-16 \*\*\*  
## month.September 25.558 2.407 10.617 <2e-16 \*\*\*  
## month.October 25.811 2.407 10.722 <2e-16 \*\*\*  
## month.November 24.803 2.407 10.303 <2e-16 \*\*\*  
## month.December 24.037 2.407 9.985 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.77 on 229 degrees of freedom  
## Multiple R-squared: 0.8407, Adjusted R-squared: 0.8324   
## F-statistic: 100.7 on 12 and 229 DF, p-value: < 2.2e-16

*Comment: The model gives the average monthly price of oil. The seasonal means model works well with the data. 83% variation in the oilprice is explained by the seasonal means model without intercept*

#Fitting seasonal means model with an intercept to oilprice data  
model3=lm(oil.price~month.)  
summary(model3)

##   
## Call:  
## lm(formula = oil.price ~ month.)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.687 -6.213 -3.312 3.243 40.477   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 25.00286 2.34938 10.642 <2e-16 \*\*\*  
## month.February -2.46886 3.36381 -0.734 0.464   
## month.March -2.12836 3.36381 -0.633 0.528   
## month.April -2.01986 3.36381 -0.600 0.549   
## month.May -1.62136 3.36381 -0.482 0.630   
## month.June -1.64086 3.36381 -0.488 0.626   
## month.July -1.32686 3.36381 -0.394 0.694   
## month.August -0.03636 3.36381 -0.011 0.991   
## month.September 0.55564 3.36381 0.165 0.869   
## month.October 0.80864 3.36381 0.240 0.810   
## month.November -0.19936 3.36381 -0.059 0.953   
## month.December -0.96586 3.36381 -0.287 0.774   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.77 on 229 degrees of freedom  
## Multiple R-squared: 0.01015, Adjusted R-squared: -0.0374   
## F-statistic: 0.2134 on 11 and 229 DF, p-value: 0.9966

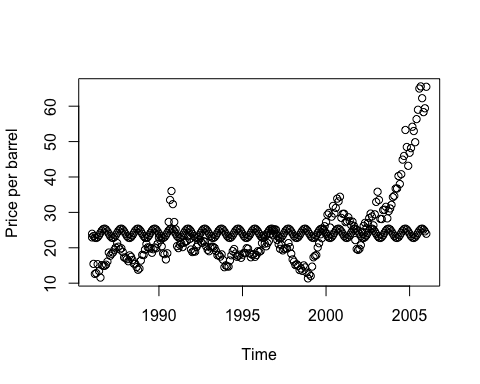
*Comment:* *The model gives the average monthly price of oil. The seasonal means model doesn't work well with the data. Not much variation in the oilprice is explained by the seasonal means model with an intercept*

#Cosine Trend Model  
har.=harmonic(oil.price,1)  
model4=lm(oil.price~har.)  
summary(model4)

##   
## Call:  
## lm(formula = oil.price ~ har.)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.287 -6.280 -3.347 3.416 41.522   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 24.0869 0.6811 35.367 <2e-16 \*\*\*  
## har.cos(2\*pi\*t) -0.1290 0.9612 -0.134 0.893   
## har.sin(2\*pi\*t) -1.3230 0.9651 -1.371 0.172   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.57 on 238 degrees of freedom  
## Multiple R-squared: 0.007908, Adjusted R-squared: -0.0004289   
## F-statistic: 0.9486 on 2 and 238 DF, p-value: 0.3888

*Comment:* *In this output, time is measured in years, with 1986 as the starting value and a frequency of 1 per year.*

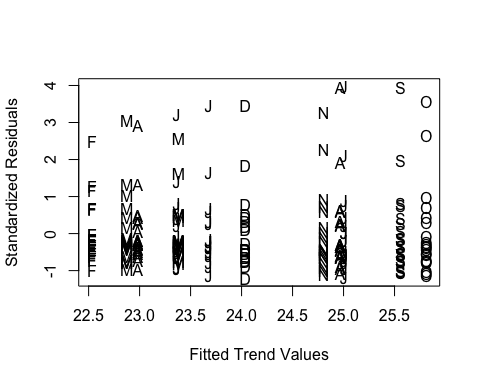
#graph of the time series values together with the fitted cosine curve  
plot(ts(fitted(model4),freq=12,start=c(1986,1)),  
 ylab='Price per barrel',type='o',  
ylim=range(c(fitted(model4),oil.price))); points(oil.price)



# ylim ensures that the y axis range fits the raw data and thefitted values

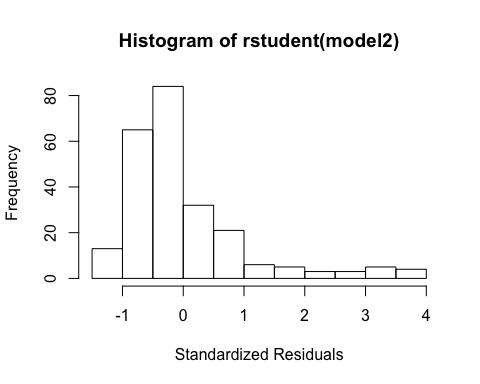
*Comment: The trend doesn't fit the data quite well, where the observations are scattered from the points predicted by the model*

#Standardized Residuals versus Fitted Values for the Temperature Seasonal Means Model  
plot(y=rstudent(model2),x=as.vector(fitted(model2)),  
 xlab='Fitted Trend Values',  
 ylab='Standardized Residuals',type='n')  
points(y=rstudent(model2),x=as.vector(fitted(model2)),  
 pch=as.vector(season(oil.price)))



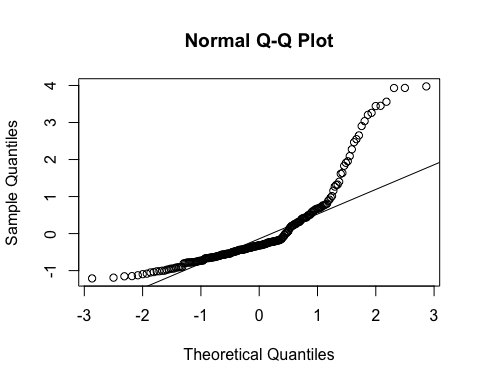
*Comment: More variation could be seen across all the months for higher residuals. Highest variation can be seen in January, April and September values.*

#Histogram of the residuals or standardized residuals.  
hist(rstudent(model2),xlab='Standardized Residuals')



*Comment: Histogram of the residuals or standardized residuals. The plot looks not symmetric and tails off at the lower ends. The plot looks skewed towards the lower end.*

qqnorm(rstudent(model2))  
qqline(rstudent(model2))



*Comment: The deviated pattern here doesn't support the assumption of a normally distributed stochastic component in this model.*

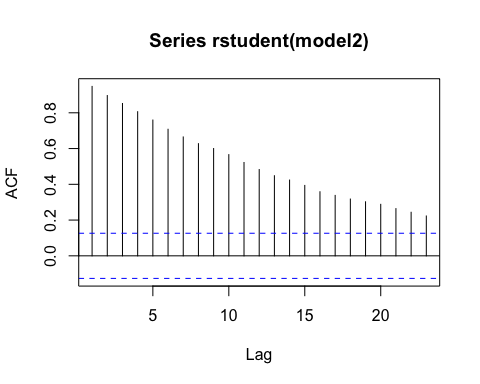
#Shapiro-Wilk test  
shapiro.test(rstudent(model2))

##   
## Shapiro-Wilk normality test  
##   
## data: rstudent(model2)  
## W = 0.78095, p-value < 2.2e-16

*Comment: The p-value is less than 0.05 therefore we can reject the null hypothesis that the stochastic component of this model is normally distributed. However, Shapiro-Wilk test works well only for a sample size < 50.*

CHAPTER 4

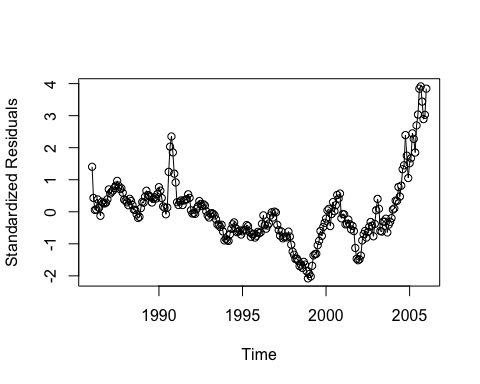
#The sample autocorrelation function of the standardized residuals  
acf(rstudent(model2))



max.lag=30

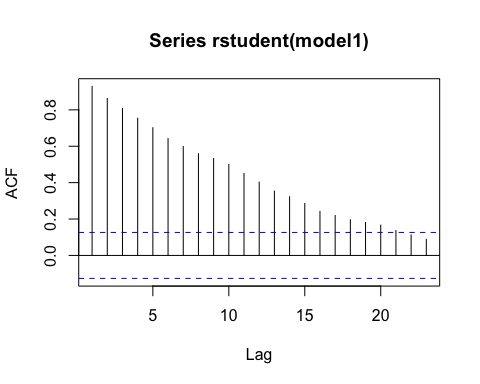
*Comment: The magnitude of the autocorrelation function decreases linearly as the number of lags, k, increases.*

#Residuals from Straight Line Fit of the oil price  
plot(y=rstudent(model1),x=as.vector(time(oil.price)),  
 ylab='Standardized Residuals',xlab='Time',type='o')



*Comment: Mostly residuals hang together. The plot does appear smooth. More variation could be seen in the end values of the series*

#Sample Autocorrelation of Residuals from Straight Line Model  
acf(rstudent(model1))

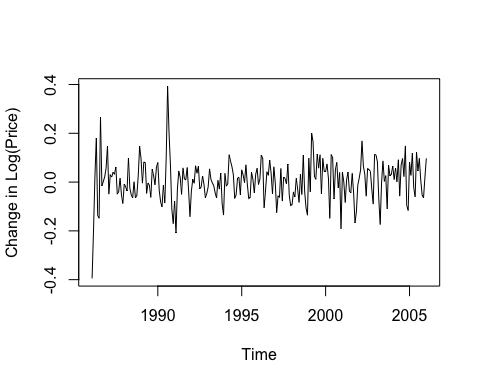


*Comment: Positive correlations seen in the series. The magnitude of the autocorrelation function decreases linearly as the number of lags, k, increases.*

CHAPTER 5

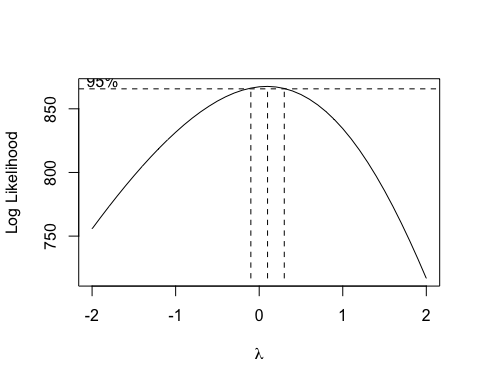
*As the series displays variation, stationary model does not seem to be reasonable*

#Stationarity through differencing  
#The Difference Series of the Logs of the Oil Price Time  
plot(diff(log(oil.price)),ylab='Change in Log(Price)',  
type='l')



*Comment: The difference series looks stationary compared to the original series. Since the data values are all positive we can go with Box cox transformation.*

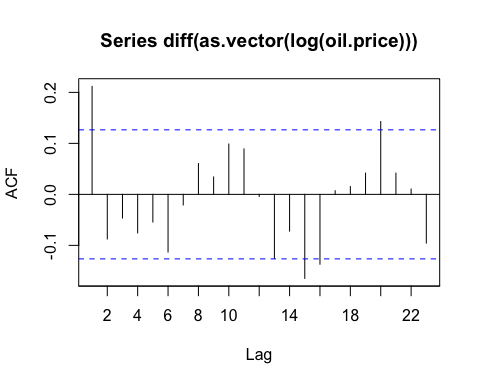
#power transformation  
BoxCox.ar((oil.price))



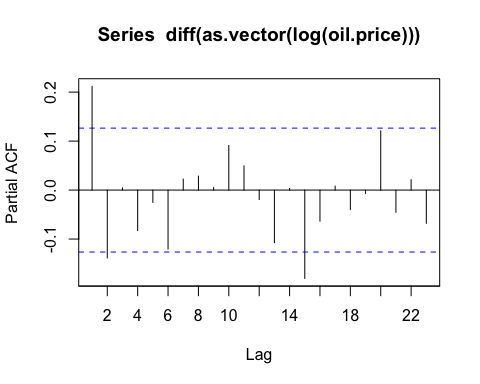
*Comment: The 95% confidence interval for lambda contains the value of lambda = 0 quite near its center and strongly suggests a logarithmic transformation (lambda = 0) for these data.*

CHAPTER 6

#Sample ACF and PACF for the Difference of the Logged Oil Price Series  
acf(diff(as.vector(log(oil.price))),xaxp=c(0,24,12))



pacf(diff(as.vector(log(oil.price))),xaxp=c(0,24,12))



*Comment: Now the pattern emerges much more clearly after differencing. ACF suggests a moving average model of order 1 (MA(1)) while PACF suggests a AR(2) model.*

#EACF table  
eacf(diff(log(oil.price)))

## AR/MA  
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13  
## 0 x o o o o o o o o o o o o o   
## 1 x x o o o o o o o o x o o o   
## 2 o x o o o o o o o o o o o o   
## 3 o x o o o o o o o o o o o o   
## 4 o x x o o o o o o o o o o o   
## 5 o x o x o o o o o o o o o o   
## 6 o x o x o o o o o o o o o o   
## 7 x x o x o o o o o o o o o o

*Comment: EACF table for the differences of the logarithms of the oil price data. This table suggests an ARMA model with p = 0 and q = 1.*

CHAPTER 7

#Estimation for the Difference of Logs of the Oil Price Series  
arima(as.vector(log(oil.price)),order=c(0,1,1),method='CSS')

##   
## Call:  
## arima(x = as.vector(log(oil.price)), order = c(0, 1, 1), method = "CSS")  
##   
## Coefficients:  
## ma1  
## 0.2731  
## s.e. 0.0681  
##   
## sigma^2 estimated as 0.006731: part log likelihood = 259.58

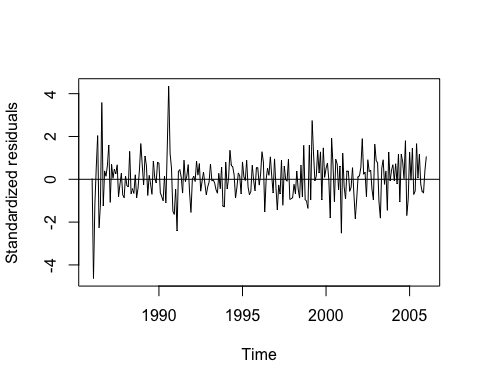
arima(as.vector(log(oil.price)),order=c(0,1,1),method='ML')

##   
## Call:  
## arima(x = as.vector(log(oil.price)), order = c(0, 1, 1), method = "ML")  
##   
## Coefficients:  
## ma1  
## 0.2956  
## s.e. 0.0693  
##   
## sigma^2 estimated as 0.006689: log likelihood = 260.29, aic = -518.58

*Comment: The sample ACF suggested an MA(1) model on the differences of the logs of the prices. Estimate of the MA(1) parameter is given here. Both the estimates are nearly equal and they are statistically significant.*

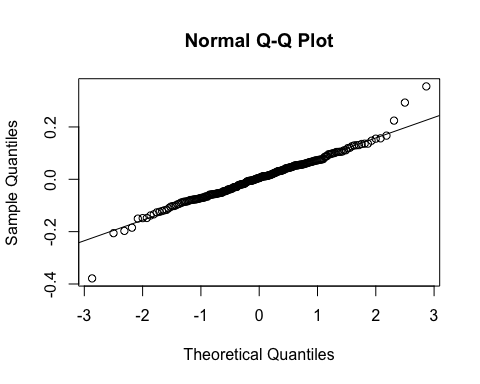
CHAPTER 8

#Standardized Residuals from Log Oil Price IMA(1,1) Model  
data(oil.price)  
m1.oil=arima(log(oil.price),order=c(0,1,1))  
plot(rstandard(m1.oil),ylab='Standardized residuals',type='l')  
abline(h=0)



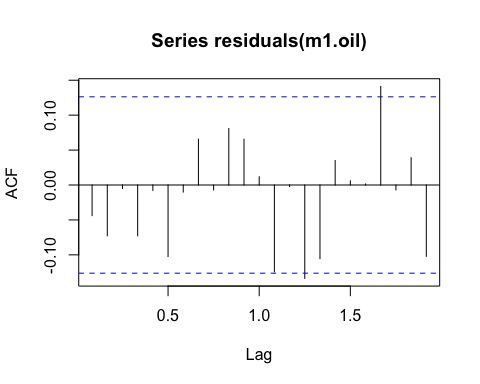
*Comment: There are at least two or three residuals early in the series with magnitudes larger than 3—very unusual in a standard normal distribution.*

#Quantile-Quantile Plot: Residuals from IMA(1,1) Model for Oil  
qqnorm(residuals(m1.oil)); qqline(residuals(m1.oil))



*Comment: The QQ plot appears normal with most of the points lying close to the line. However, the plot shows a little deviation from the normal line in the top.*

acf(residuals(m1.oil))



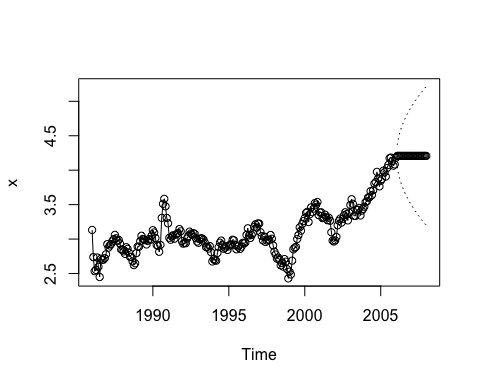
*Comment: Lag 1.3 and 1.7 are significant. Other lags show little evidence for autocorrelation of the residual terms of this model. The series can be modeled as a IMA(1,1) model with uncorrelated, normal residuals or error terms.*

CHAPTER 9

modeloilforecast=arima(log(oil.price),order=c(0,1,1))  
plot(modeloilforecast,n.ahead=24)$pred

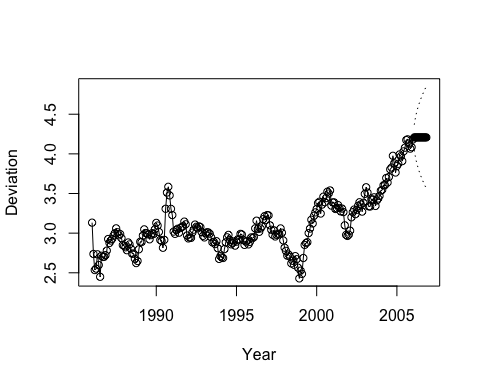
## Jan Feb Mar Apr May Jun Jul Aug  
## 2006 4.20755 4.20755 4.20755 4.20755 4.20755 4.20755 4.20755  
## 2007 4.20755 4.20755 4.20755 4.20755 4.20755 4.20755 4.20755 4.20755  
## 2008 4.20755   
## Sep Oct Nov Dec  
## 2006 4.20755 4.20755 4.20755 4.20755  
## 2007 4.20755 4.20755 4.20755 4.20755  
## 2008

abline(h=0)



*Comment: The forecasts are quite constant from forecast 2006 onwards*

plot(modeloilforecast,n.ahead=10,ylab='Deviation',xlab='Year',pch=19)  
abline(h=coef(modeloilforecast)[names(coef(modeloilforecast))=='intercept'])



*Comment: The forecasts plotted at the end in dark circles, quickly settled down to a value as the model does not contain any pattern or strong autocorrelation*

**6.35** The data file named deere3 contains 57 consecutive measurements recorded from a complex machine tool at Deere & Co. The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced.

**(a)** Display the time series plot of this series and comment on its appearance. Would a stationary model be appropriate here?

**(b)** Display the sample ACF and PACF for this series and select tentative orders for an ARMA model for the series.

*Solution:*

library(TSA)

## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1 2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-17. For overview type 'help("mgcv-package")'.

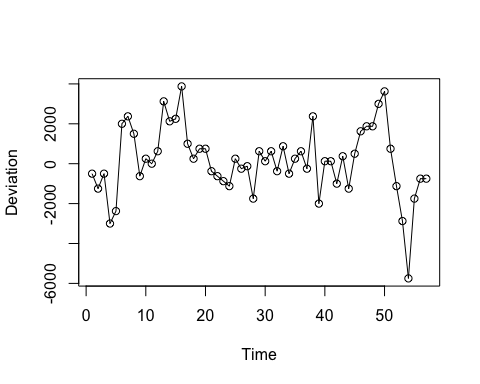
## Loading required package: tseries

##   
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':  
##   
## acf, arima

## The following object is masked from 'package:utils':  
##   
## tar

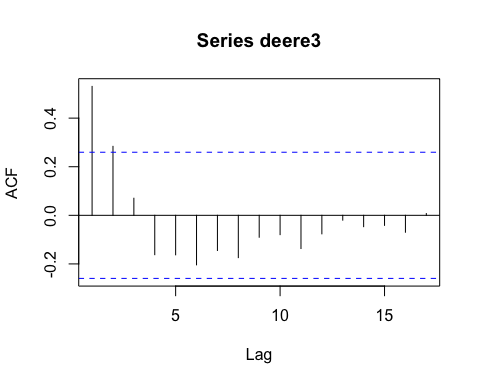
data(deere3)  
plot(deere3, type='o',ylab='Deviation')



*Comment*: The plot looks stationary except for the last few observations

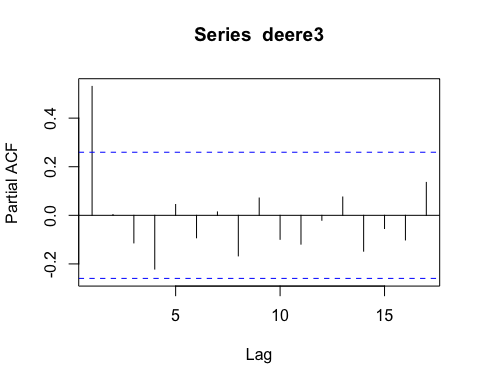
1. Display the sample ACF and PACF for this series and select tentative orders for an ARMA model for the series.

acf(deere3)



*Comment:* The plot strongly suggests an AR(1) model for this series.

pacf(deere3)



*Comment:* The plot strongly suggests an AR(1) model for this series.

7.28 The data file named deere3 contains 57 consecutive values from a complex machine tool at Deere & Co. The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced.

**(a)** Estimate the parameters of an AR(1) model for this series.

**(b)** Estimate the parameters of an AR(2) model for this series and compare the results with those in part (a).

*Solution:*

1. Estimate the parameters of an AR(1) model for this series.

library(TSA)

## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1 2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-17. For overview type 'help("mgcv-package")'.

## Loading required package: tseries

##   
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':  
##   
## acf, arima

## The following object is masked from 'package:utils':  
##   
## tar

data(deere3)  
arima(deere3,order=c(1,0,0))

##   
## Call:  
## arima(x = deere3, order = c(1, 0, 0))  
##   
## Coefficients:  
## ar1 intercept  
## 0.5255 124.3832  
## s.e. 0.1108 394.2067  
##   
## sigma^2 estimated as 2069355: log likelihood = -495.51, aic = 995.02

*Comment:* The AR(1) coefficient is significantly different from 0.

1. Estimate the parameters of an AR(2) model for this series and compare the results with those in part (a).

arima(deere3,order=c(2,0,0))

##   
## Call:  
## arima(x = deere3, order = c(2, 0, 0))  
##   
## Coefficients:  
## ar1 ar2 intercept  
## 0.5211 0.0083 123.2979  
## s.e. 0.1310 0.1315 397.6134  
##   
## sigma^2 estimated as 2069208: log likelihood = -495.51, aic = 997.01

*Comment:* The AR(2) coefficient is not significantly different from 0 so the AR(1) model can be appropriate.

8.10 The data file named deere3 contains 57 consecutive values from a complex machine tool at Deere & Co. The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced. Diagnose the fit of an AR(1) model for these data in terms of the tests discussed in this chapter.

*Solution:*

library(TSA)

## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1 2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-17. For overview type 'help("mgcv-package")'.

## Loading required package: tseries

##   
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':  
##   
## acf, arima

## The following object is masked from 'package:utils':  
##   
## tar

data(deere3)  
model=arima(deere3,order=c(1,0,0));  
model

##   
## Call:  
## arima(x = deere3, order = c(1, 0, 0))  
##   
## Coefficients:  
## ar1 intercept  
## 0.5255 124.3832  
## s.e. 0.1108 394.2067  
##   
## sigma^2 estimated as 2069355: log likelihood = -495.51, aic = 995.02

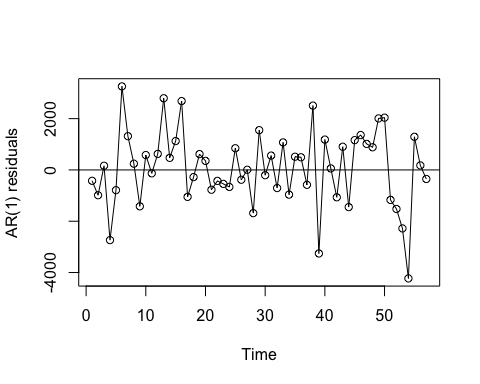
*Comment:* The AR(1) term is significantly statistically. However, we could remove the intercept from the model. The model can be fit by removing mean or intercept term

model1=arima(deere3,order=c(1,0,0),include.mean=F);  
model1

##   
## Call:  
## arima(x = deere3, order = c(1, 0, 0), include.mean = F)  
##   
## Coefficients:  
## ar1  
## 0.5291  
## s.e. 0.1103  
##   
## sigma^2 estimated as 2072748: log likelihood = -495.56, aic = 993.12

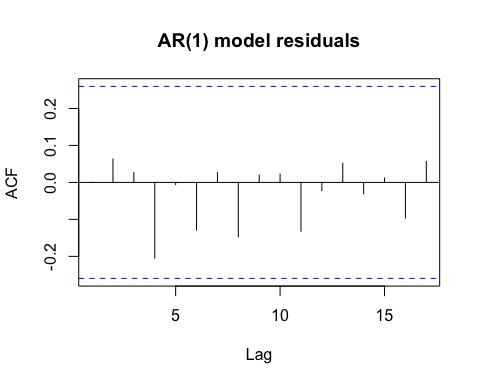
*Comment:* There is a slight changed in AR(1) coefficient estimate. AIC of model is pretty worse for this model.

residual=residuals(model1);  
plot(residual,ylab='AR(1) residuals',type='o');  
abline(h=0)



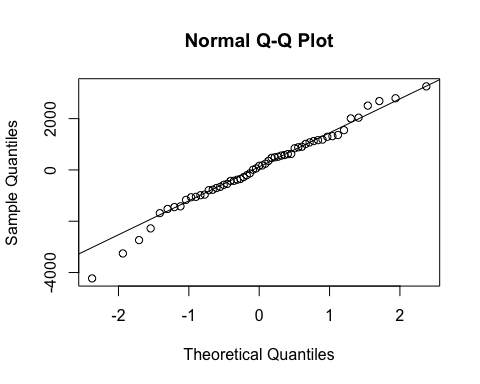
*Comment:* Residual looks little random

acf(residual,main='AR(1) model residuals')



*Comment:* There is no or little evidence of autocorrelation of the residual terms of this model

qqnorm(residual)  
qqline(residual)



shapiro.test(residual)

##   
## Shapiro-Wilk normality test  
##   
## data: residual  
## W = 0.98289, p-value = 0.5966

*Comment:* The QQ plot shows a little deviation from the normal line in the bottom. However, the shapiro-wilk test is clear. Normality assumption is satisfied and cannot be rejected. The series can be modeled as a AR(1) model with no intercept and uncorrelated, normal residuals or error terms.

9.21 The data file named deere3 contains 57 consecutive values from a complex machine tool process at Deere & Co. The values given are deviations from a tar- get value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced.

**(a)** Using an AR(1) model for this series, forecast the next ten values.

**(b)** Plot the series, the forecasts, and 95% forecast limits, and interpret the results.

*Solution:*

1. Using an AR(1) model for this series, forecast the next ten values.

library(TSA)

## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1 2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-17. For overview type 'help("mgcv-package")'.

## Loading required package: tseries

##   
## Attaching package: 'TSA'

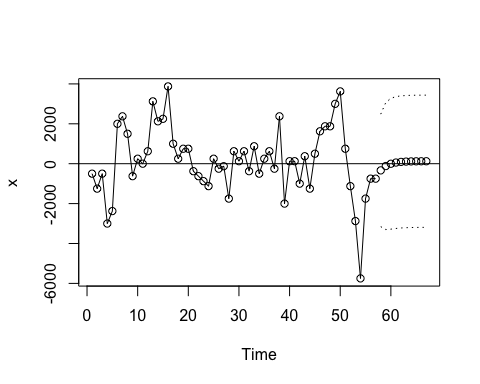
## The following objects are masked from 'package:stats':  
##   
## acf, arima

## The following object is masked from 'package:utils':  
##   
## tar

data(deere3)  
model=arima(deere3,order=c(1,0,0))  
plot(model,n.ahead=10)$pred

## Time Series:  
## Start = 58   
## End = 67   
## Frequency = 1   
## [1] -335.145928 -117.120772 -2.538388 57.679997 89.327566  
## [6] 105.959839 114.700873 119.294695 121.708962 122.977772

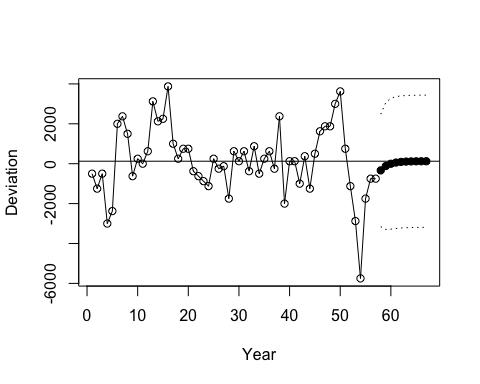
abline(h=0)



*Comment:* The forecasts are quite constant from forecast 8 onwards

1. Plot the series, the forecasts, and 95% forecast limits, and interpret the results.

plot(model,n.ahead=10,ylab='Deviation',xlab='Year',pch=19)  
abline(h=coef(model)[names(coef(model))=='intercept'])



*Comment:* The forecasts plotted at the end in dark circles, quickly settled down to the mean as the model does not contain any pattern or autocorrelation