PSTAT 174 Project

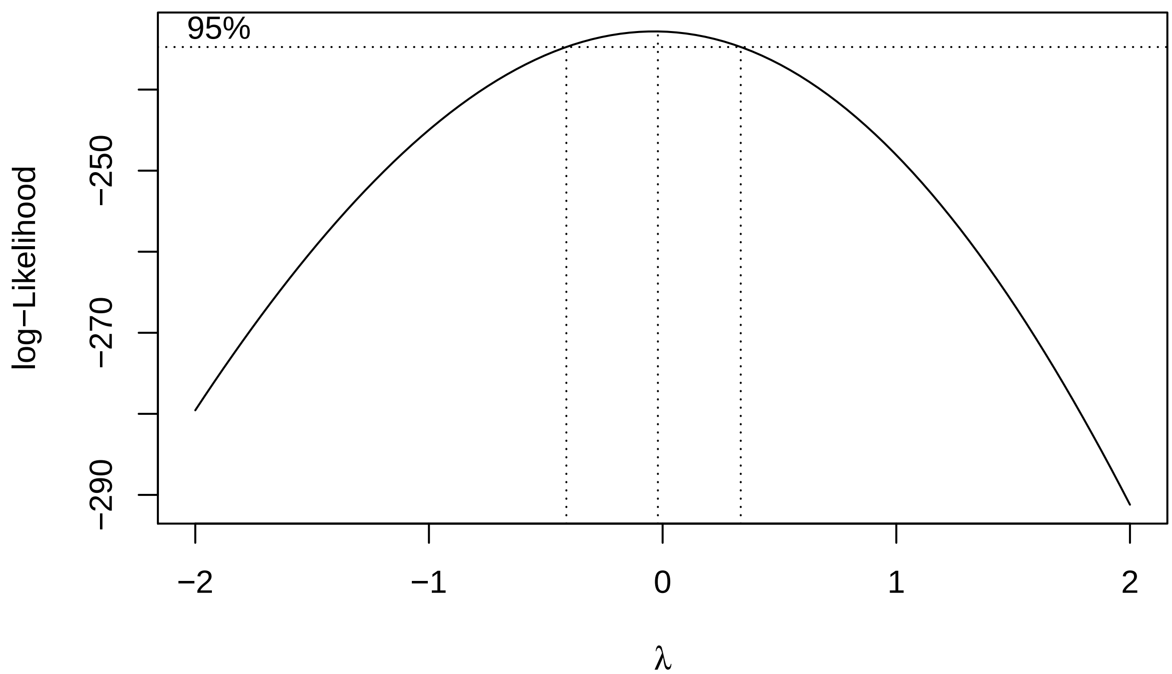
Group: Theta

Winter 2018

```r
library(forecast)
library(qpcR)
library(MASS)
library(car)
```

Creating and transforming data set.

```r
sp500=read.csv("C:/Users/kebro/Desktop/Pstat 174/sp500monthly.csv",header=T)
sp500.close=ts((sp500$Close),frequency = 12,start=c(2000,1))#creating time series data-set
sp500.close.bc=boxcox(sp500.close ~ as.numeric(1:length(sp500.close)))
```



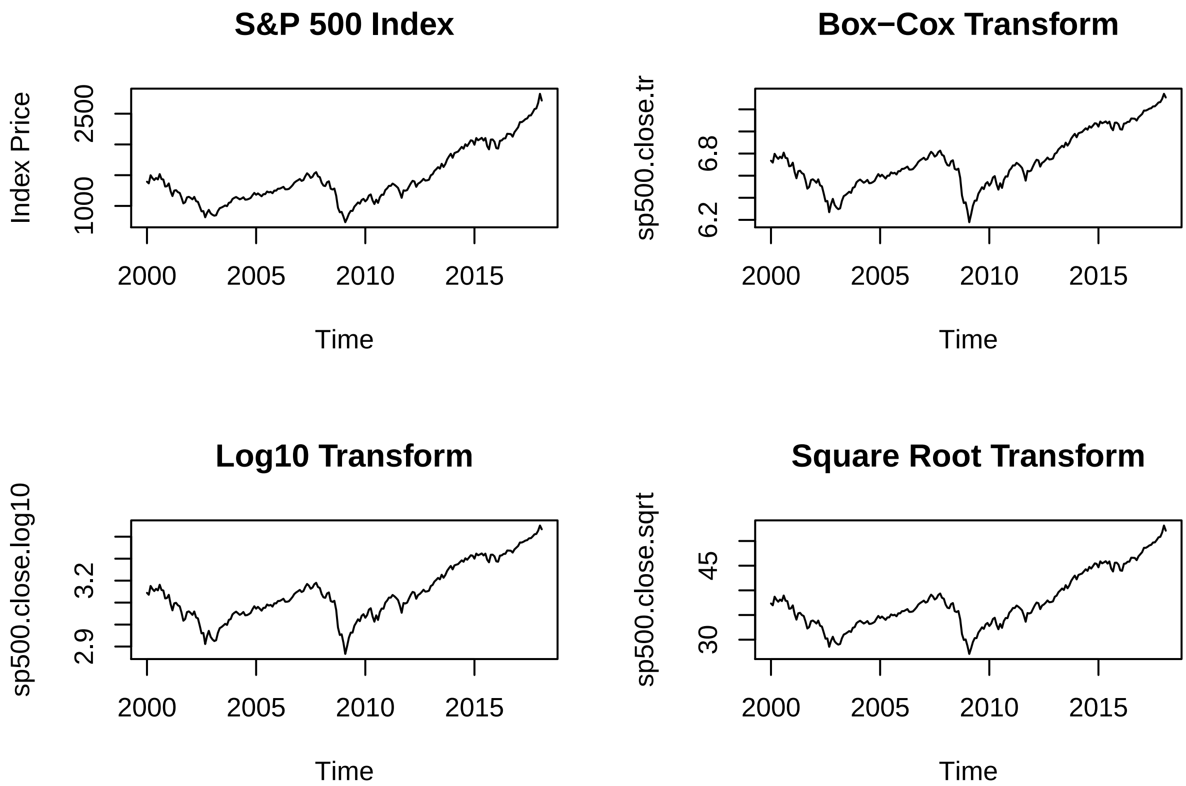

```r
sp500.close.log10=log10(sp500.close)#log10 transform data
sp500.close.sqrt=sqrt(sp500.close)#sqrt transformed data
lambda=sp500.close.bc$x[which(sp500.close.bc$y==max(sp500.close.bc$y))]
lambda
```

```
## [1] -0.02020202
```

```r
sp500.close.tr=(1/lambda)*(sp500.close^lambda-1)#box-cox transformed data
```

Plotting each transform

```
op=par(mfrow=c(2,2))
plot.ts(sp500.close,main="S&P 500 Index",xlab="Time",ylab="Index Price")
ts.plot(sp500.close.tr,main="Box-Cox Transform")
ts.plot(sp500.close.log10,main="Log10 Transform")
ts.plot(sp500.close.sqrt,main="Square Root Transform")
```

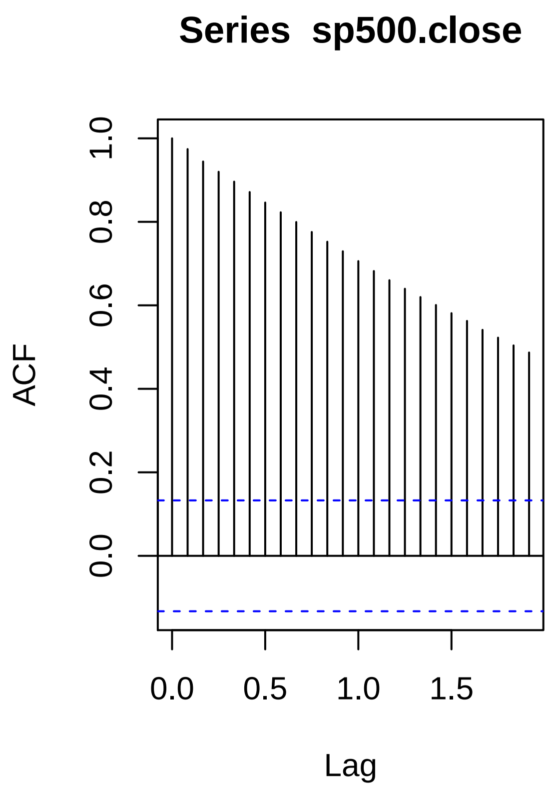


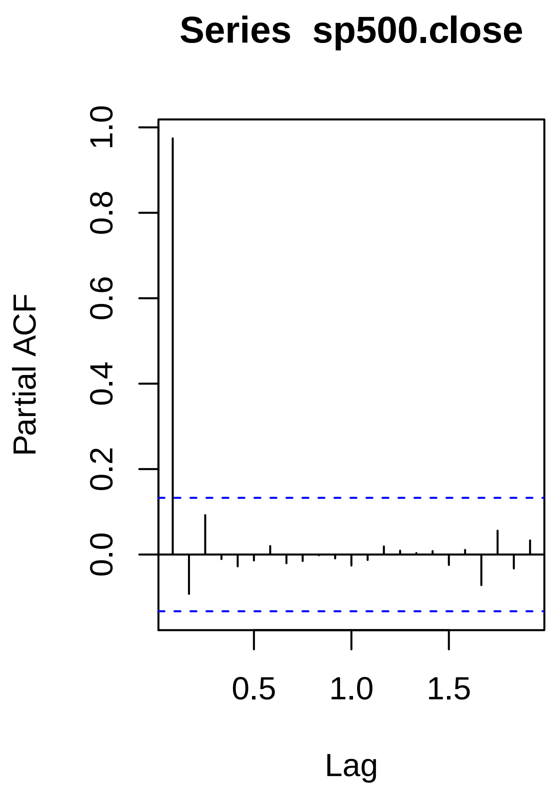

```
par(op)
```

No distinct diference between each data set. Transforming the data has no real effect. Thus we choose original dataset for all further analysis.

ACF and PACF of original data set

```
op=par(mfrow=c(1,2))
acf(sp500.close)
pacf(sp500.close)
```






```r
par(op)
```

Differencing the data set once to remove trend.

```r
sp500.diff1=diff(sp500.close,1)
var(sp500.diff1)
```

```
## [1] 2994.366
```

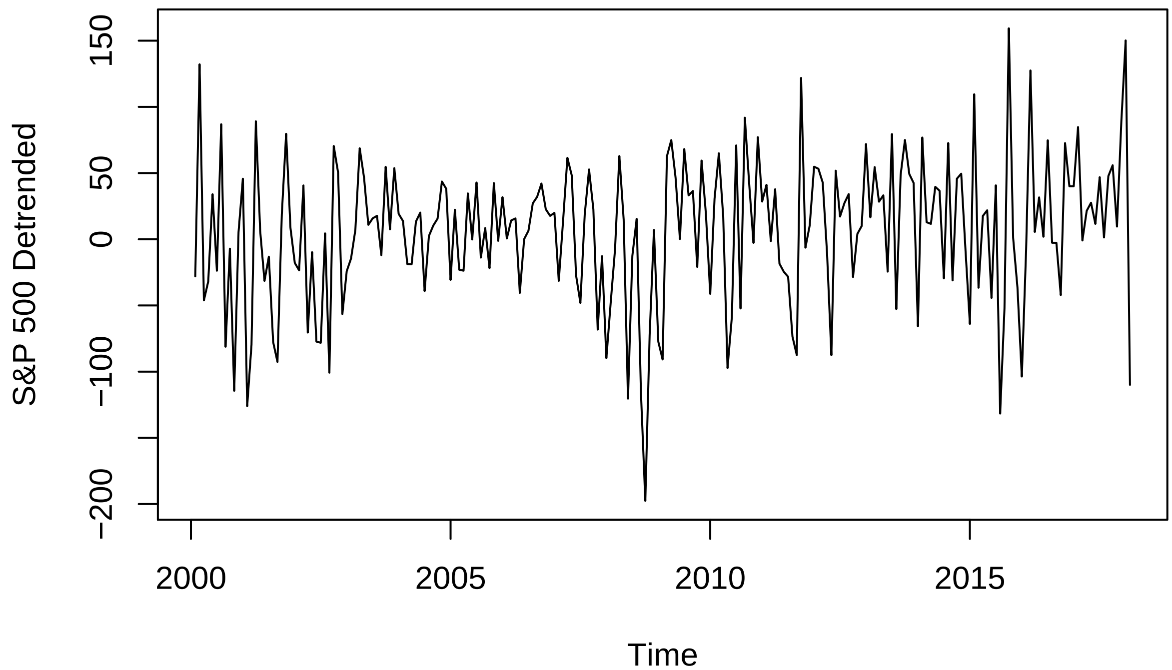
```r
sp500.diff2=diff(sp500.close,2)
var(sp500.diff2)
```

```
## [1] 6198.822
```

```r
sp500.diff3=diff(sp500.close,3)
var(sp500.diff3)
```

```
## [1] 9138.741
```

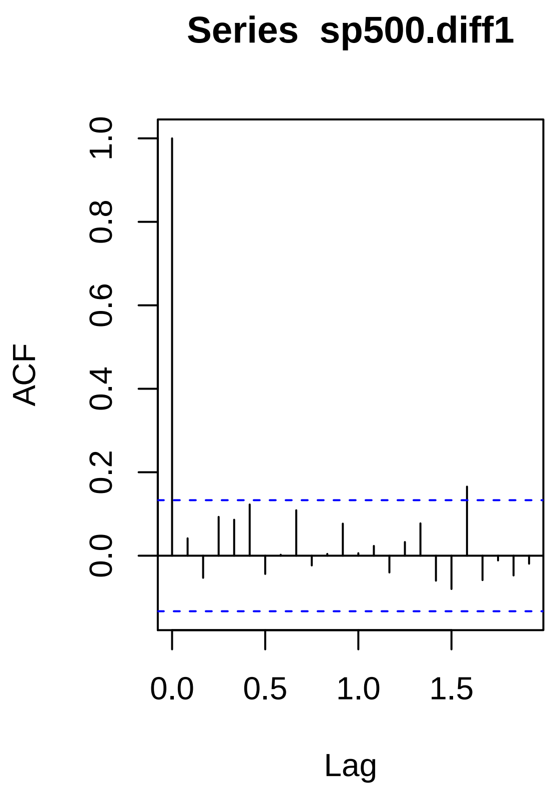
```r
ts.plot(sp500.diff1,ylab="S&P 500 Detrended")
```

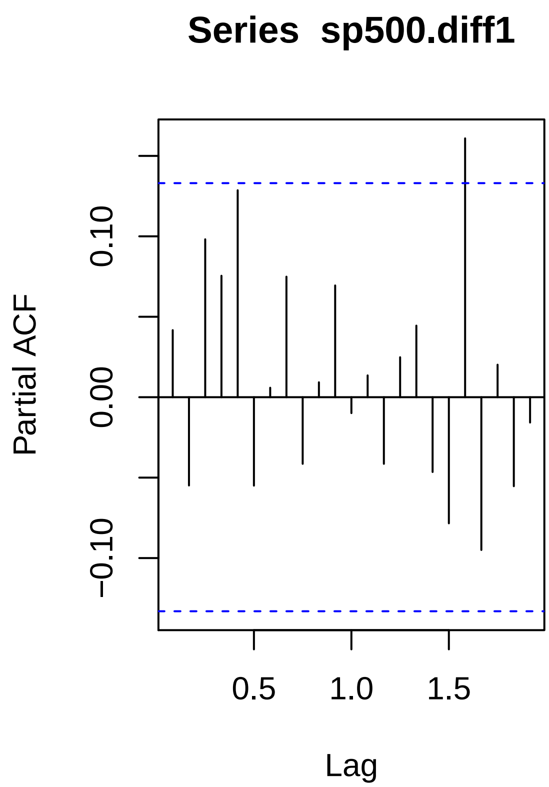


Differencing once results in least amount of variance in the data set. We choose a d=1 to be our differencing parameter.

ACF and PACF of detrended data

```
op=par(mfrow=c(1,2))
acf(sp500.diff1)
pacf(sp500.diff1)
```






```r
par(op)
```

Creating a matrix of values from which we brute force our way to an ideal ARIMA model

```r
sp500.AICcs=matrix(NA, nr = 6, nc = 6)
dimnames(sp500.AICcs) = list(p=0:5, q=0:5)
```

```r
for(p in 0:5){
  for(q in 0:5){
    sp500.AICcs[p+1,q+1] = AICc(arima(sp500.diff1, order = c(p,0,q), method="ML",optim.control = list(ma
  }}
```

```r
sp500.AICcs
```

```
##    q
## p          0        1        2        3        4        5
##   0 2355.809 2357.409 2358.604 2357.926 2358.830 2358.492
##   1 2357.463 2353.403 2355.477 2357.565 2359.625 2360.259
##   2 2358.775 2355.477 2357.565 2359.615 2357.534 2357.382
##   3 2358.493 2357.765 2359.734 2358.594 2359.632 2360.491
##   4 2359.317 2359.796 2357.571 2359.281 2360.620 2357.140
##   5 2357.715 2359.322 2359.965 2358.509 2360.372 2361.021
```

```r
sp500.AICcs==min(sp500.AICcs)
```

```
##    q
## p       0     1     2     3     4     5
##   0 FALSE FALSE FALSE FALSE FALSE FALSE
```

```
##   1 FALSE  TRUE FALSE FALSE FALSE FALSE
##   2 FALSE FALSE FALSE FALSE FALSE FALSE
##   3 FALSE FALSE FALSE FALSE FALSE FALSE
##   4 FALSE FALSE FALSE FALSE FALSE FALSE
##   5 FALSE FALSE FALSE FALSE FALSE FALSE
```

Find ARIMA(1,0,1) to be ideal for this method

Using auto.arima function to calculate another potential model

```
sp500.arima=auto.arima(sp500.close,max.p = 20,max.q=20)
summary(sp500.arima)
```

```
## Series: sp500.close
## ARIMA(1,2,1)(1,0,2)[12]
##
## Coefficients:
##          ar1      ma1      sar1    sma1    sma2
##       0.0160  -0.9718  -0.4611  0.4483  0.0427
## s.e.  0.0753   0.0202   0.0290  0.0337  0.0640
##
## sigma^2 estimated as 3046:  log likelihood=-1171.75
## AIC=2355.5   AICc=2355.9   BIC=2375.75
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE     MASE
## Training set 3.856443 54.29739 42.15126 0.2271881 3.186486 0.214681
##                      ACF1
## Training set -0.002085993
```

Arive at ARIMA(1,2,1)(1,0,2)[12]. Despite having a slightly larger AICc we choose this model as it accounts for seasonality.

Residual Testing of Chosen Model

```
sp500.arima.residuals=sp500.arima$residuals
Box.test(sp500.arima.residuals,lag=20,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  sp500.arima.residuals
## X-squared = 21.814, df = 20, p-value = 0.3507
```

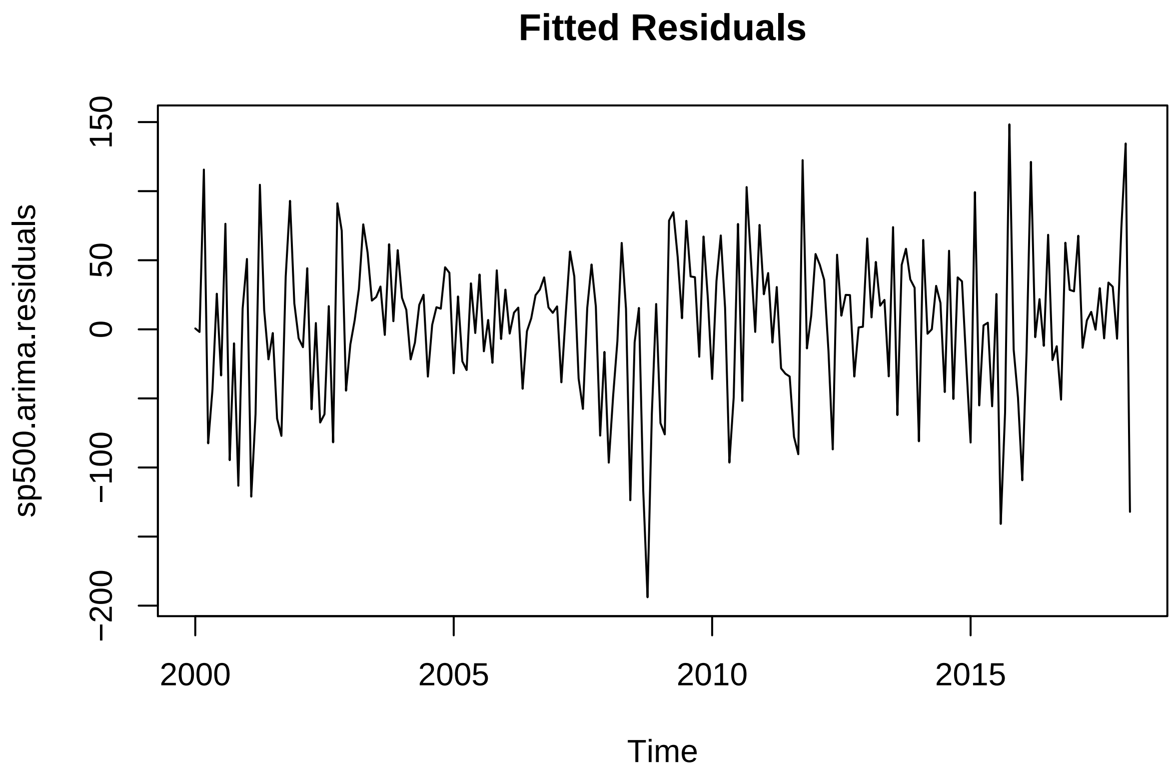
```
shapiro.test(sp500.arima.residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sp500.arima.residuals
## W = 0.98573, p-value = 0.0274
```

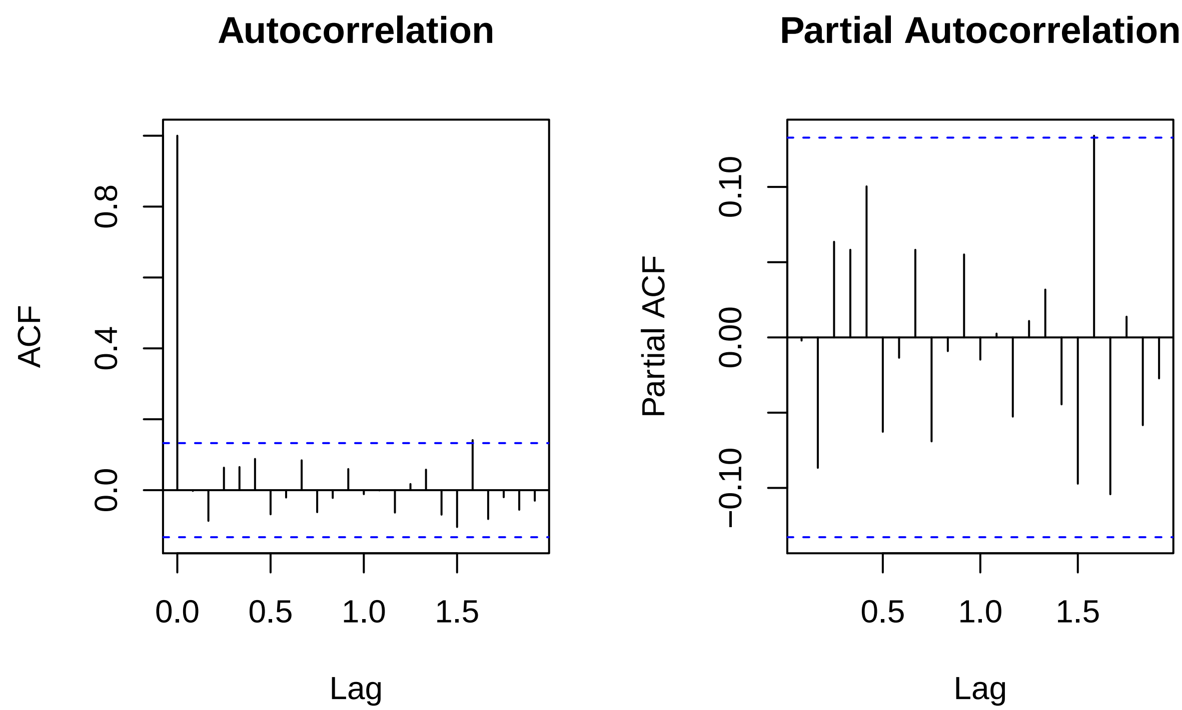
Fails shapiro wilk test due to outlier

Residual diagnostic plots imply potential normality with outliers. Histogram is overlayed with a gausian curve, qq-normailty plot overlayed with 95% confidence intervals.

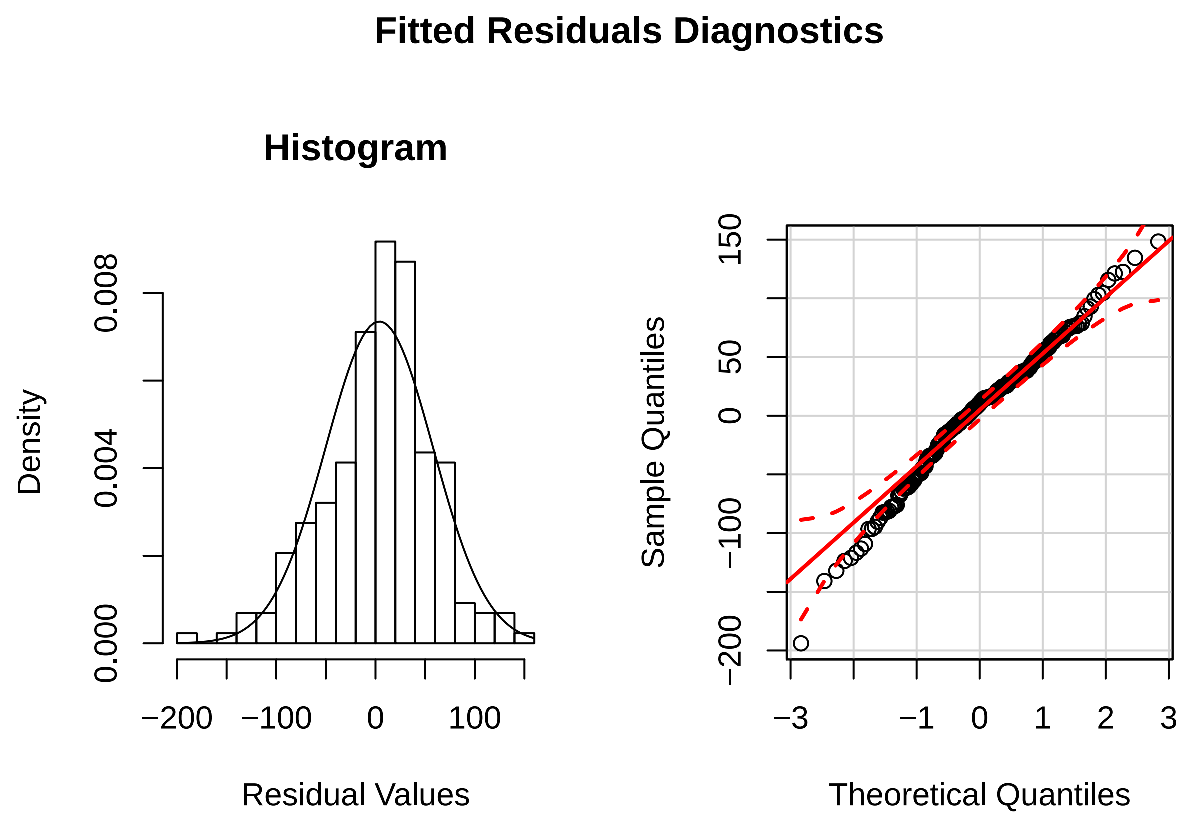
```
ts.plot(sp500.arima.residuals,main="Fitted Residuals")
```




```
op=par(mfrow=c(1,2),oma=c(0,0,2,0))
acf(sp500.arima.residuals,main="Autocorrelation")
pacf(sp500.arima.residuals,main="Partial Autocorrelation")
```




```
hist(sp500.arima.residuals,main="Histogram",breaks=20,xlab="Residual Values",probability = T)
curve(dnorm(x,mean=mean(sp500.arima.residuals),sd=sd(sp500.arima.residuals)),add=T)
qqPlot(sp500.arima.residuals,xlab="Theoretical Quantiles",ylab="Sample Quantiles")
title("Fitted Residuals Diagnostics",outer=T)
```




```r
par(op)
```

Histogram shows mild skew and significant kurtosis, several points are outside the qq-plots conf-int.

Removing single outlier with lowest value and rerunning tests.

```r
sp500.arima.residuals.no_outlier=sp500.arima$residuals[sp500.arima$residuals>min(sp500.arima$residuals)]
Box.test(sp500.arima.residuals.no_outlier,lag=20,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  sp500.arima.residuals.no_outlier
## X-squared = 17.315, df = 20, p-value = 0.6324
```

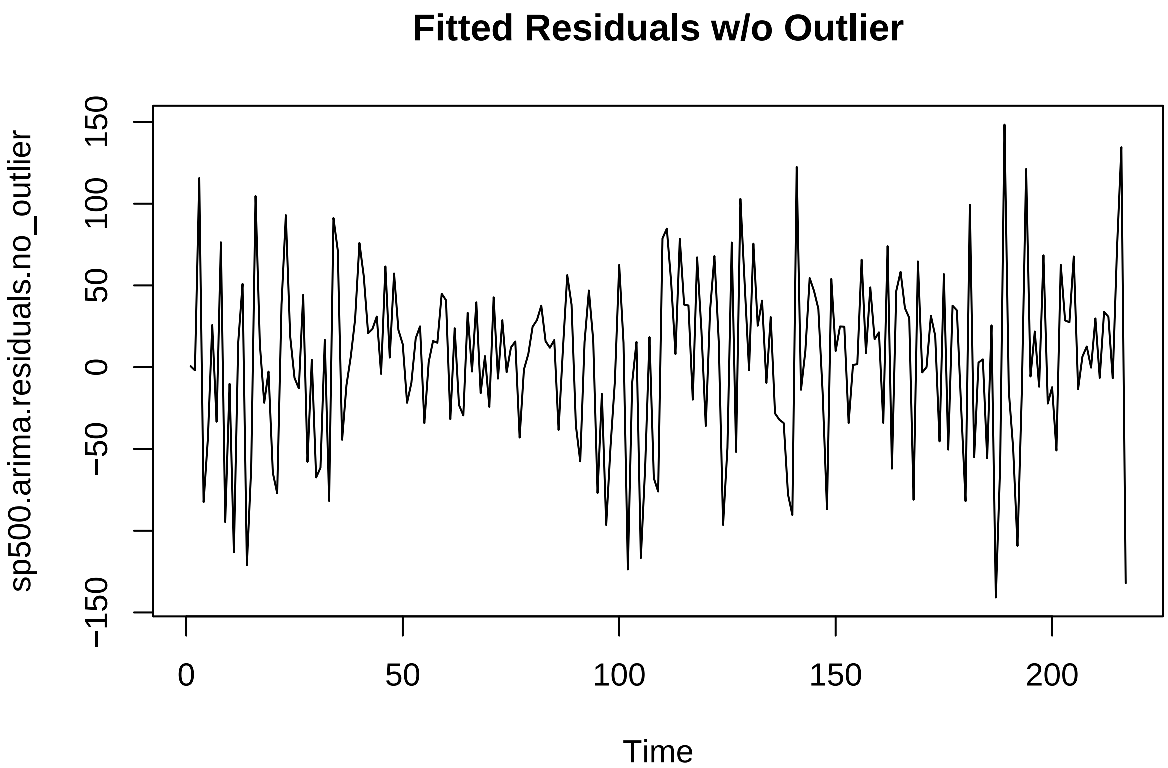
```r
shapiro.test(sp500.arima.residuals.no_outlier)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sp500.arima.residuals.no_outlier
## W = 0.98922, p-value = 0.1035
```

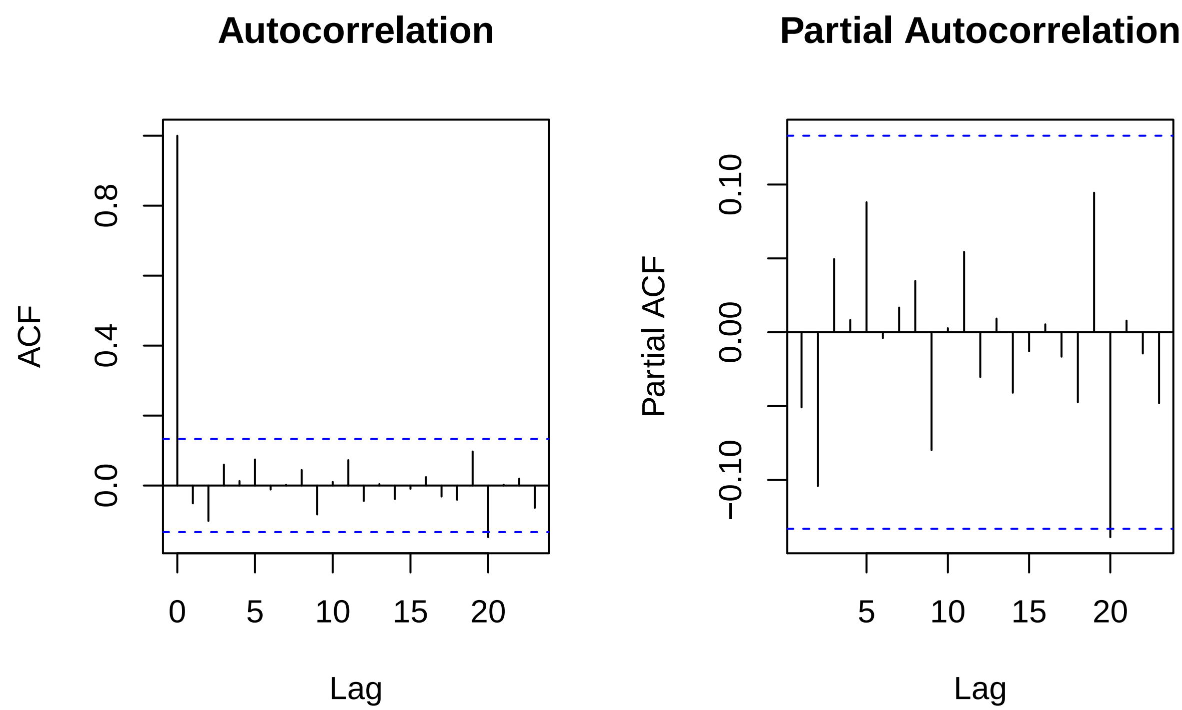
Normality assumption confirmed when lowest value residual is removed

Diagnostic plots w/o most extreme outlier.

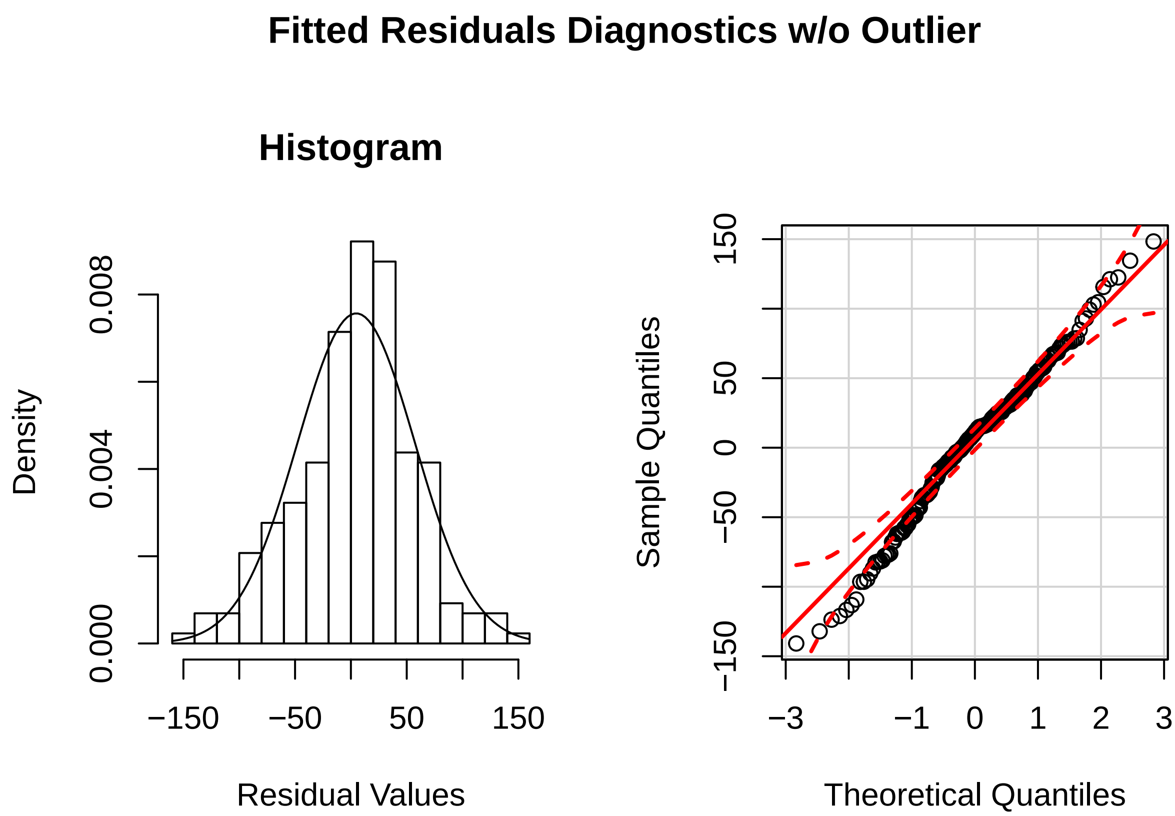
```r
ts.plot(sp500.arima.residuals.no_outlier,main="Fitted Residuals w/o Outlier")
```




```r
op=par(mfrow=c(1,2),oma=c(0,0,2,0))
acf(sp500.arima.residuals.no_outlier,main="Autocorrelation")
pacf(sp500.arima.residuals.no_outlier,main="Partial Autocorrelation")
```




```
hist(sp500.arima.residuals.no_outlier,main="Histogram",
     breaks=20,xlab="Residual Values",probability = T)
curve(dnorm(x,mean=mean(sp500.arima.residuals.no_outlier),
            sd=sd(sp500.arima.residuals.no_outlier)),add=T)
qqPlot(sp500.arima.residuals.no_outlier,xlab="Theoretical Quantiles",ylab="Sample Quantiles")
title("Fitted Residuals Diagnostics w/o Outlier",outer=T)
```




```r
par(op)
```

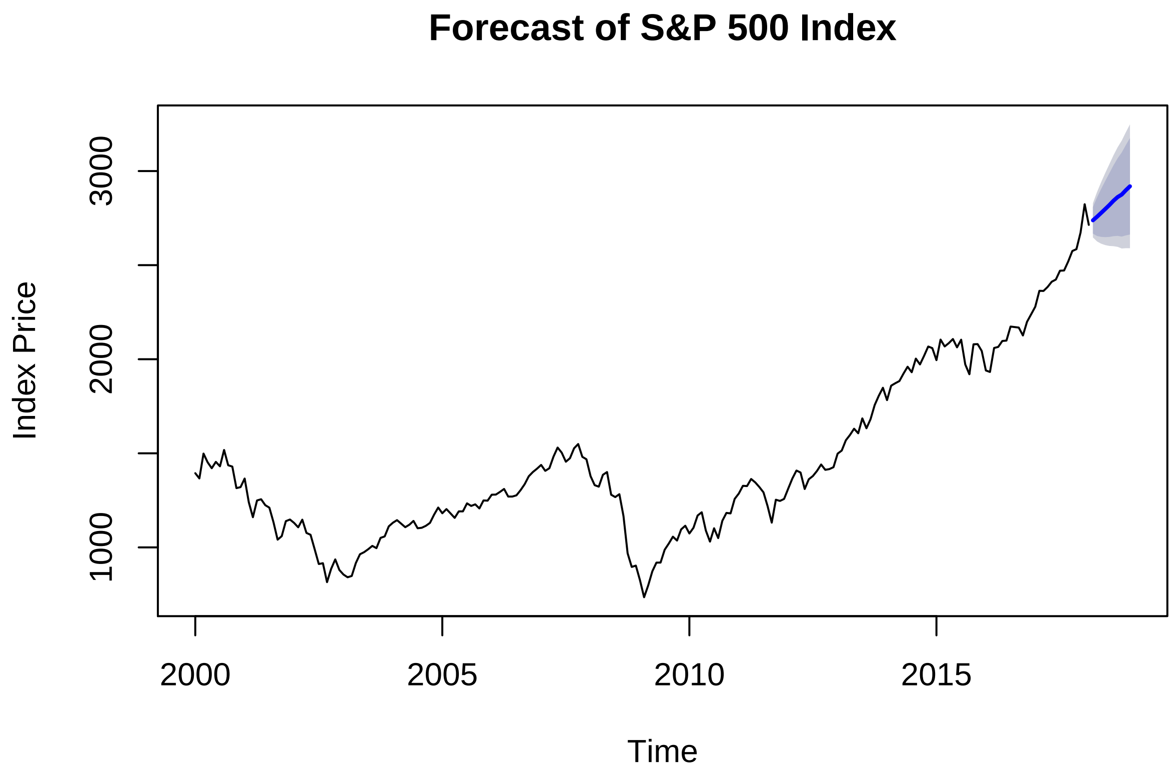
Histogram still shows kurtosis, qq-plot shows stronger normality despite a couple of outliers.

Forecast of S&P 500 with an 80% and 95% confidence interval. values of highs and lows for each level given on a monthly basis.

```r
sp500.forecast=forecast(sp500.arima,h=10,level=c(.8,.9))
sp500.forecast
```

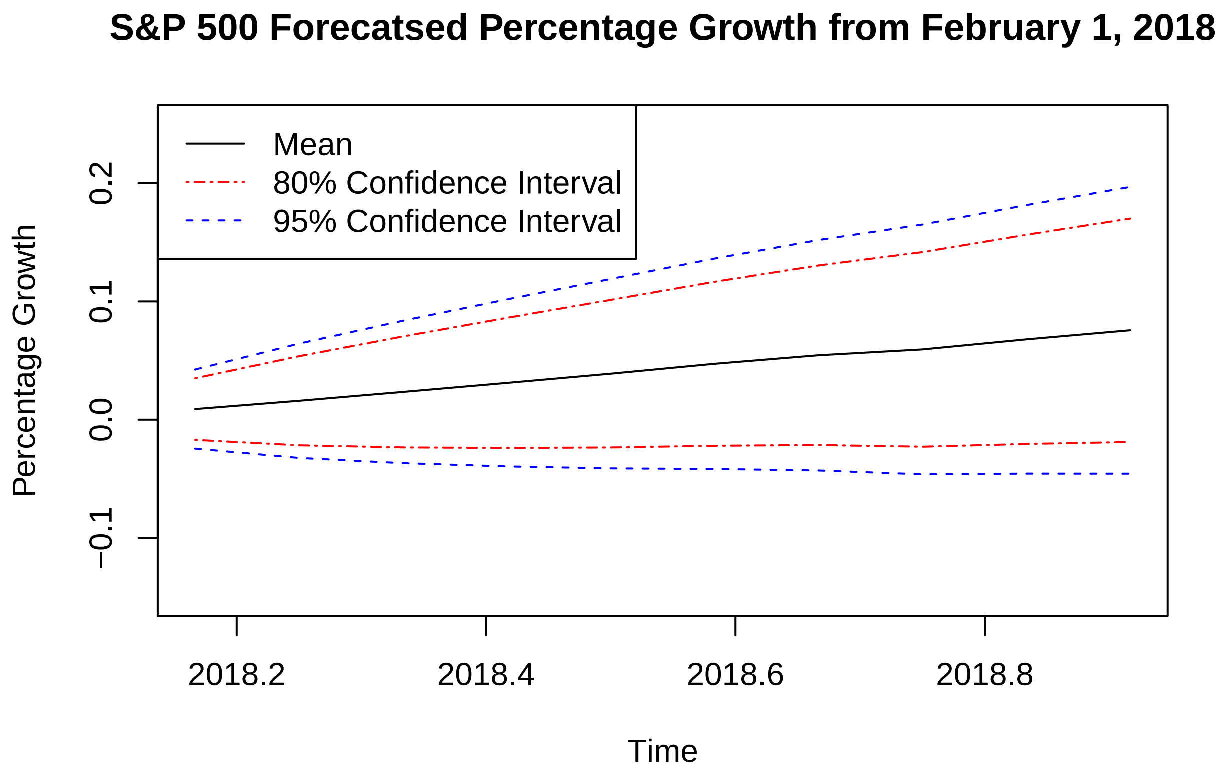
```
##          Point Forecast    Lo 80    Hi 80    Lo 90    Hi 90
## Mar 2018       2738.201 2667.471 2808.931 2647.420 2828.982
## Apr 2018       2757.227 2654.962 2859.493 2625.971 2888.484
## May 2018       2777.522 2650.163 2904.882 2614.058 2940.987
## Jun 2018       2798.212 2648.897 2947.527 2606.568 2989.856
## Jul 2018       2819.419 2650.037 2988.800 2602.020 3036.817
## Aug 2018       2842.197 2654.011 3030.382 2600.663 3083.730
## Sep 2018       2861.703 2655.609 3067.796 2597.184 3126.221
## Oct 2018       2875.124 2651.785 3098.464 2588.471 3161.778
## Nov 2018       2898.184 2658.101 3138.268 2590.040 3206.328
## Dec 2018       2919.041 2662.601 3175.481 2589.904 3248.178
```

```r
plot(sp500.forecast,ylab="Index Price",xlab="Time",main="Forecast of S&P 500 Index")
```

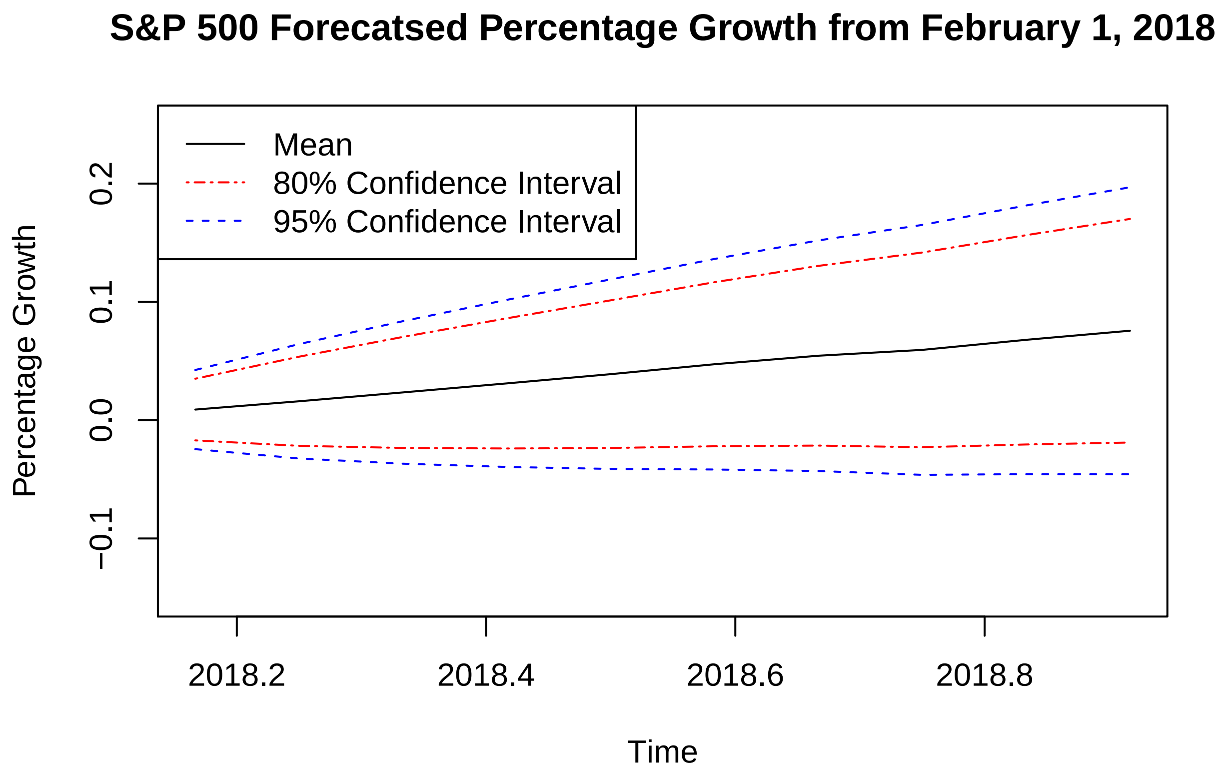


Forecast of Percentage Growth from February

```
plot(sp500.forecast$mean/sp500.close[218]-1,ylim=c(-.15,.25),
     main="S&P 500 Forecatsed Percentage Growth from February 1, 2018",col="black",
     ylab="Percentage Growth")
lines(sp500.forecast$lower[,1]/sp500.close[218]-1,col="red",lty=6)
lines(sp500.forecast$upper[,1]/sp500.close[218]-1,col="red",lty=6)
lines(sp500.forecast$lower[,2]/sp500.close[218]-1,col="blue",lty=2)
lines(sp500.forecast$upper[,2]/sp500.close[218]-1,col="blue",lty=2)
legend("topleft",legend=c("Mean","80% Confidence Interval","95% Confidence Interval"),lty=c(1,6,2),col=
```




```r
plot(sp500.forecast$mean/sp500.close[218]-1,ylim=c(-.15,.25),
     main="S&P 500 Forecatsed Percentage Growth from February 1, 2018",col="black",
     ylab="Percentage Growth")
lines(sp500.forecast$lower[,1]/sp500.close[218]-1,col="red",lty=6)
lines(sp500.forecast$upper[,1]/sp500.close[218]-1,col="red",lty=6)
lines(sp500.forecast$lower[,2]/sp500.close[218]-1,col="blue",lty=2)
lines(sp500.forecast$upper[,2]/sp500.close[218]-1,col="blue",lty=2)
legend("topleft",legend=c("Mean","80% Confidence Interval","95% Confidence Interval"),lty=c(1,6,2),col=c
```


S&P 500 Forecatsed Percentage Growth from February 1, 2018
Mean
80% Confidence Interval
95% Confidence Interval
Percentage Growth
0.2
0.1
0.0
−0.1
2018.2
2018.4
2018.6
2018.8
Time