# STAT509-001-HW9-Xinye Xu

## Q1 (c)

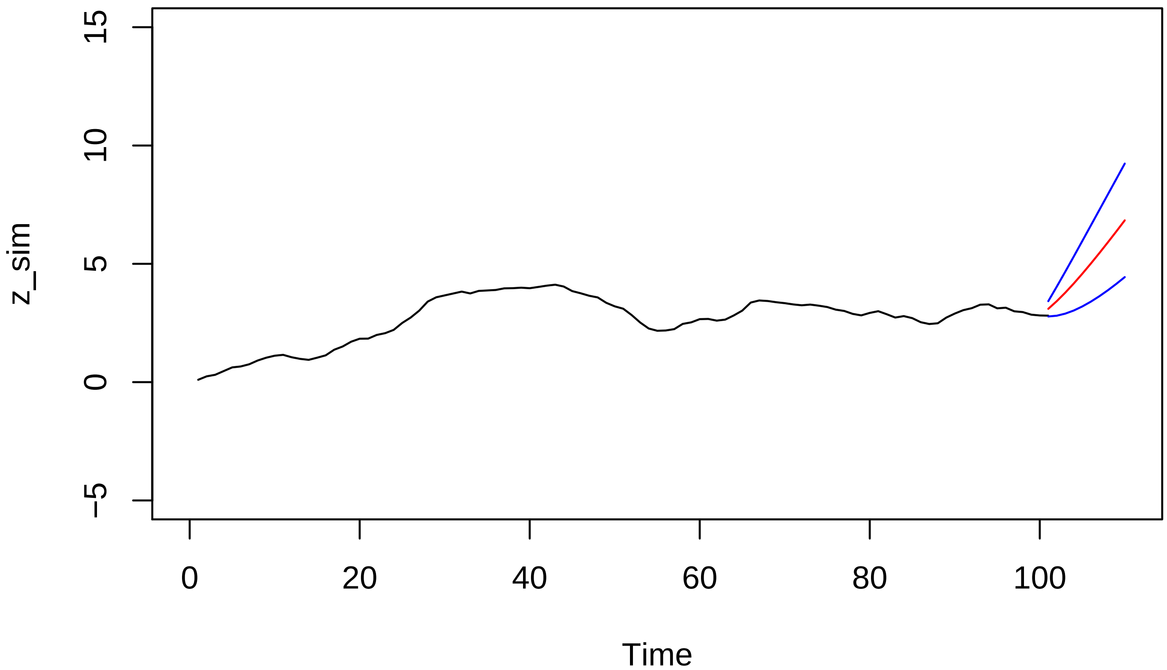
Based on your results in (b), carry out a simulation of the process Z1, Z2,. . . , Z100 and generate best linear predictions for Z101, Z102,. . . , Z110 along with 95% confidence intervals. Show a plot that contains the simulated process, the predicted values, and the confidence intervals on the predicted values. (Given $\alpha_0 = 0.1$, $\alpha_1 = 0.8$, $\sigma = 0.1$)

```r
# Simulate from an ARIMA Model: arima.sim
alpha0 = 0.1
alpha1 = 0.8
sigma = 0.1
z_sim = arima.sim(n = 100, list(order = c(1,1,0), ar = alpha1), sd = sigma) + alpha0

gamma_z = function(n,k,alpha1 = 0.8,sigma=0.1){
  part1 = sigma^2/((1-alpha1^2)*(alpha1-1)^2)
  part2 =  alpha1^(n+1+k) - alpha1^(1+k) +
    alpha1^(n+1) - alpha1 - n*alpha1^2 + n
  return(part1*part2)
}

MSPE = function(n,k,alpha1 = 0.8,sigma=0.1){
  gamma_z(n+k,0) - (gamma_z(n,k)^2)/(gamma_z(n,0))
}

# 100 predictions
n = 100
z_predict = rep(NA, 10)
se_predict = rep(NA, 10)
for (k in c(1:10)){
  z_predict[k] = (n+k)*alpha0/(1-alpha1) +
    (gamma_z(n,k)/gamma_z(n,0)) * (z_sim[n]-n*alpha0/(1-alpha1))
  se_predict[k] = sqrt(MSPE(n,k))
}
plot(z_sim, xlim = c(0,110), ylim = c(-5,15))
lines(seq(101,110), z_predict, col = "red")
lines(seq(101,110), z_predict + 1.96*se_predict, col = "blue")
lines(seq(101,110), z_predict - 1.96*se_predict, col = "blue")
```

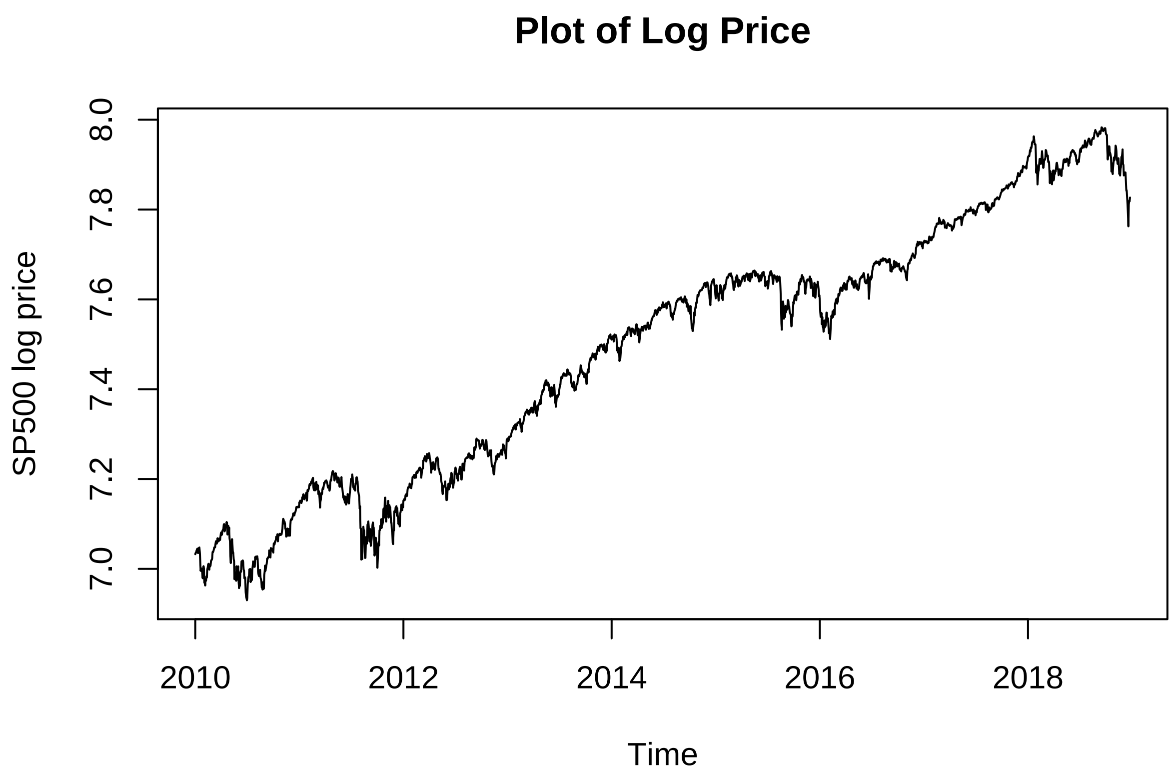



**Q2**

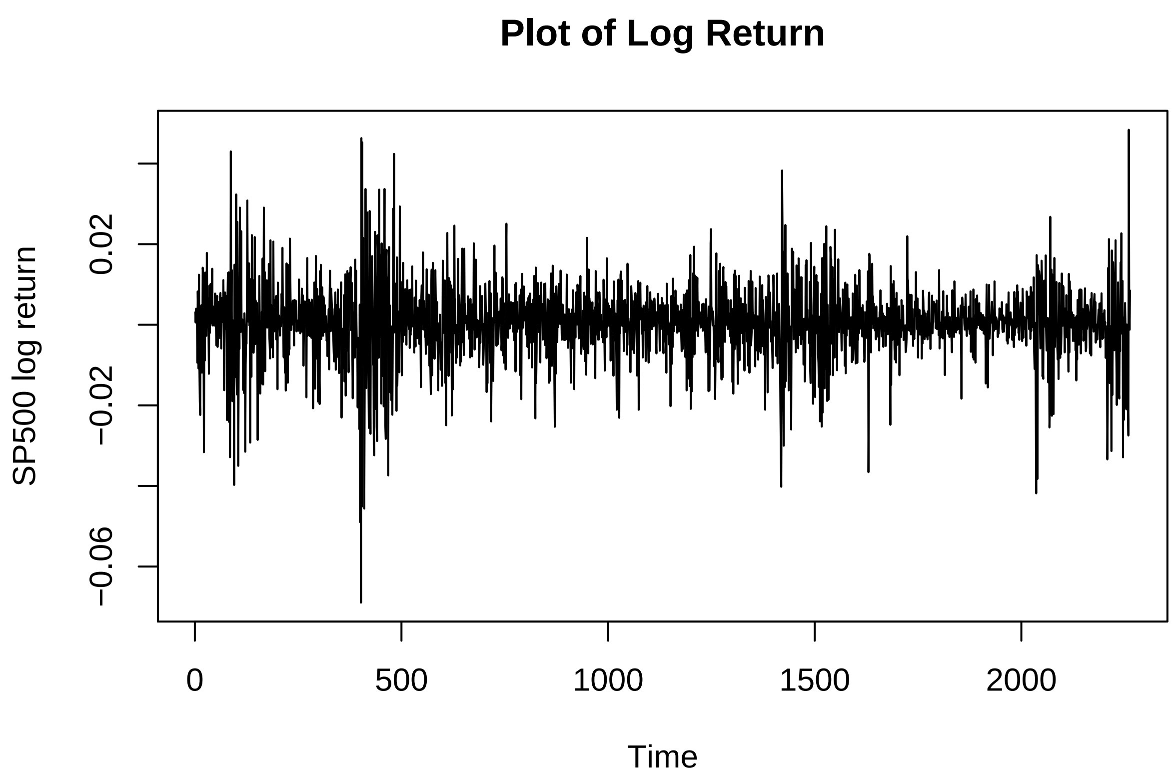
(a) For SP500 daily data from Jan 1 2010 to Dec 31 2018 (in data folder in Files) plot the logarithm of the adjusted closing price and the log-returns. Also, show plots of the autocorrelation functions of the log-price and log-returns, and provide a discussion on what these plots indicate.

The plot of Log-price clealy shows that log price is not stationary and it has a positive incresing trend.It is also supported by the ACF test, that all sticks are ouside of the confidence interval. So log price is not stationary. Log return looks like stationary in the plot, and the ACF plot confirms our conclusion as no stick beyong the CI. From these plots, we can know taht log price has very high autocorrelations for most lags, and it is non-stationary. But for log return, all of the autocorrelations up to 25 of can be reagarded as zero.

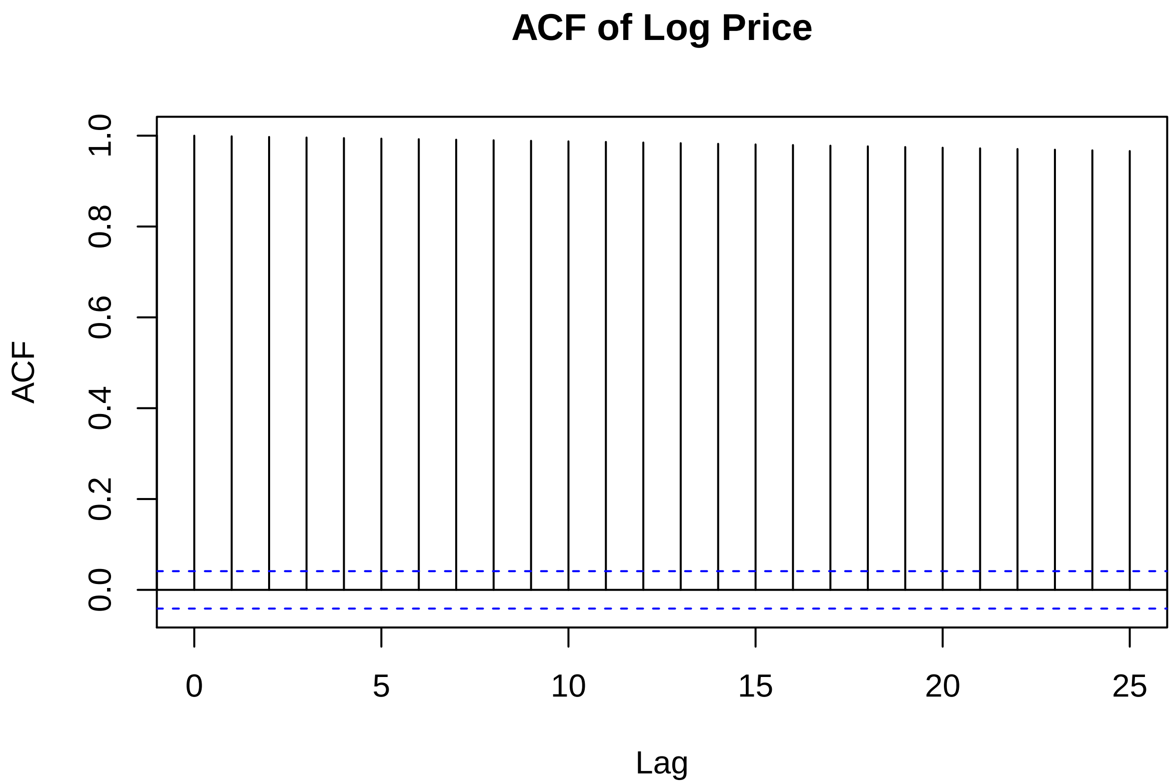
```
data <- read.csv("SP500_Jan01_2010_Dec31_2018.csv", header = T)
log_price <-log(data$Adj.Close)
log_price_ts<-ts(log_price, start = c(2010,1,4), frequency = 252)
log_return <- diff(log_price)
plot(log_price_ts,xlab="Time",ylab="SP500 log price",main="Plot of Log Price",type="l")
```



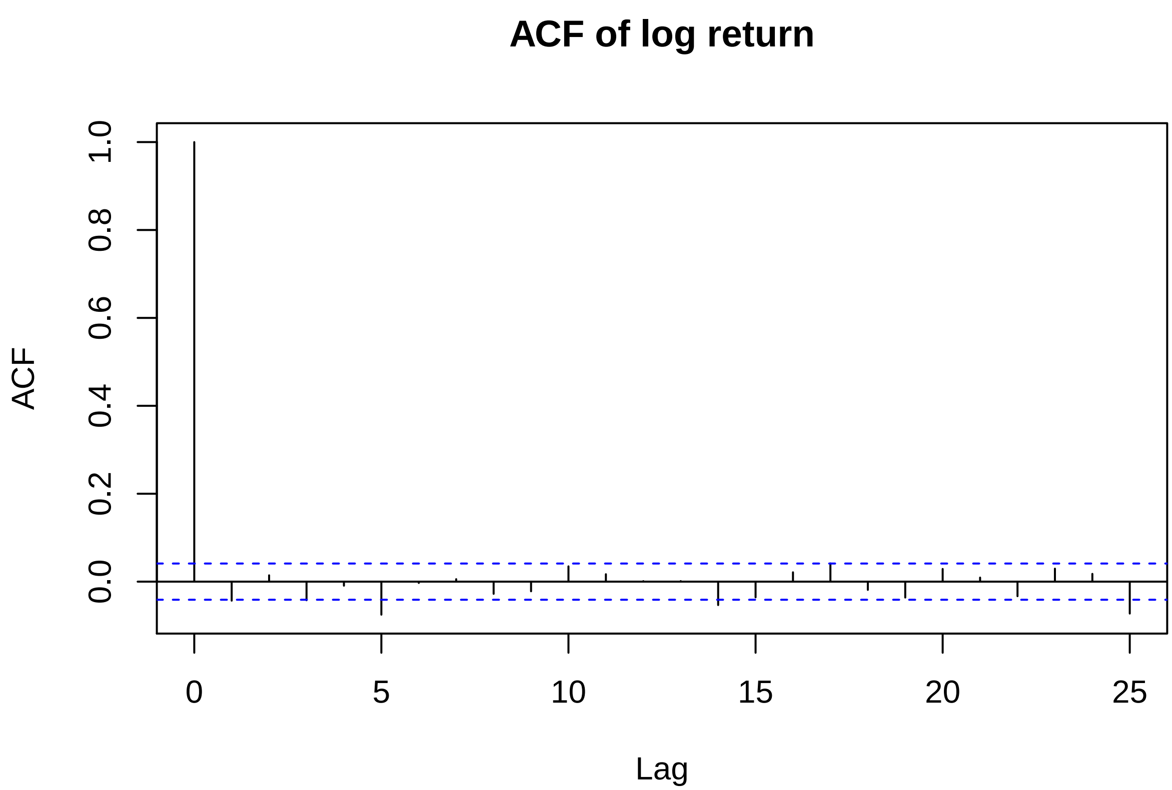

```
plot(log_return,xlab="Time",ylab="SP500 log return",main="Plot of Log Return",type="l")
```




```
acf(as.vector(log_price_ts),lag=25,main='ACF of Log Price')
```




```r
acf(as.vector(log_return),lag=25,main='ACF of log return')
```




(b) Carry out a fitting of an ARIMA model to the logarithm of the adjusted closing price for SP500, utilizing the AIC criteria for choosing model. Provide a thorough discussion on the final model selected and a full set of diagnostics.

Notice the diagostics plots from command tsdiag, standardized residual plot does not seem to have a pattern. The ACF sticks for all lags are in the CI, but from Ljung-Box tests, some p-values are significant after lag 4, suggesting the reject of autocorrelations are all zero for residuals lags up to 25. Also, the QQ plot indicaes the residuals have a heavier tail compared with normal distribution.

```r
library(forecast)
```

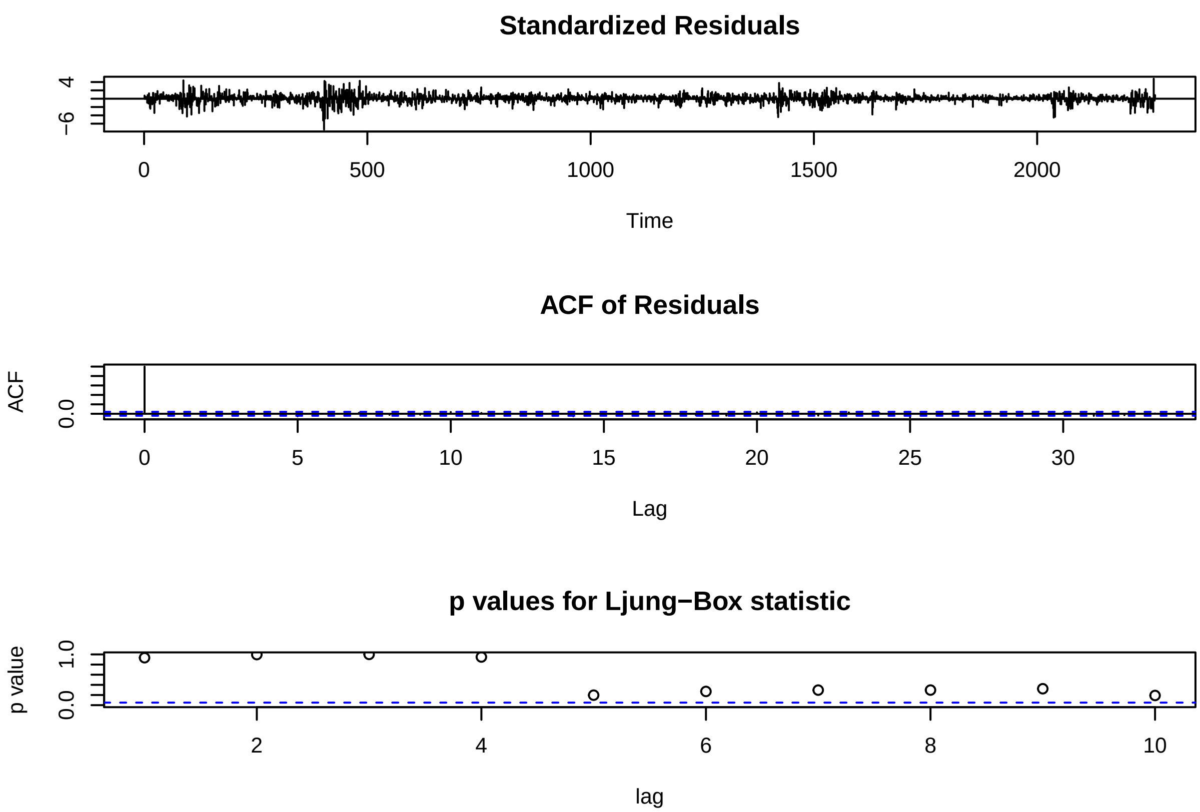
```
## Warning: package 'forecast' was built under R version 3.4.4
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'zone/tz/2018i.
## 1.0/zoneinfo/America/Detroit'
```

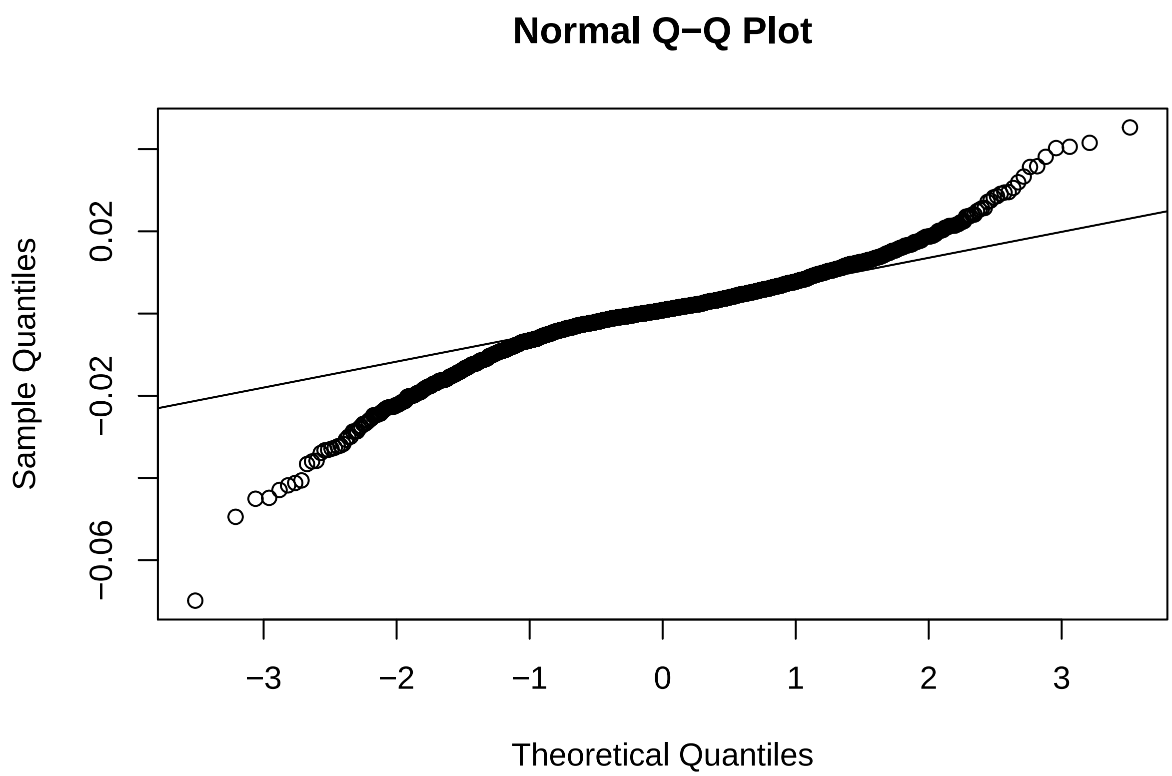
```r
auto.arima(log_price,max.P=5,max.Q=5,ic="aic")
```

```
## Series: log_price
## ARIMA(1,1,3) with drift
##
## Coefficients:
##          ar1      ma1     ma2      ma3  drift
##       0.7820  -0.8286  0.0438  -0.0486  4e-04
## s.e.  0.1388   0.1397  0.0281   0.0228  2e-04
##
## sigma^2 estimated as 8.912e-05:  log likelihood=7343.49
## AIC=-14674.98   AICc=-14674.95   BIC=-14640.64
```

```r
fit = arima(log_price,order=c(1,1,3))
tsdiag(fit)
```

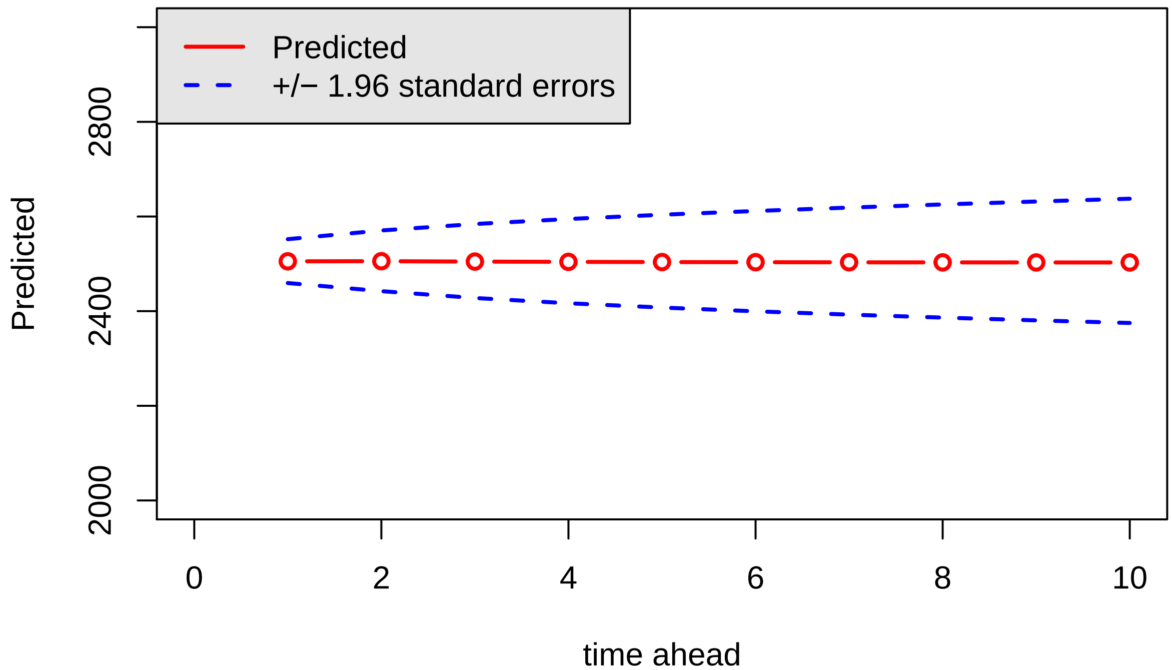



```r
qqnorm(fit$resid)
qqline(fit$resid)
```

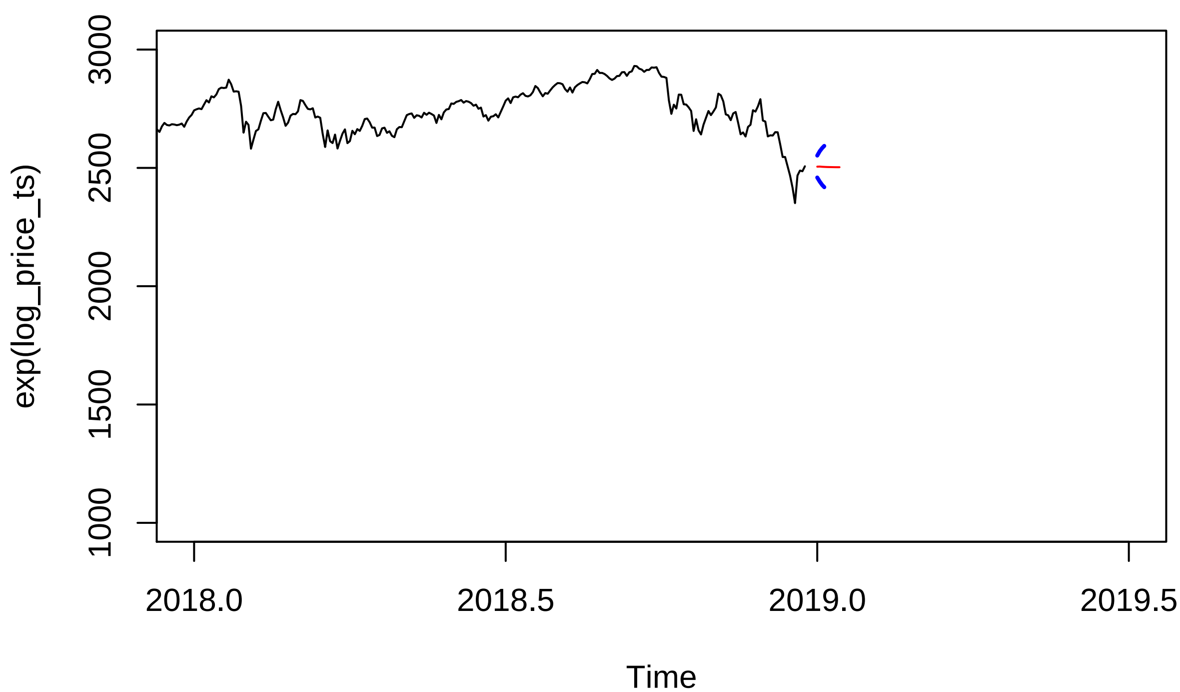



(c) Based on your work in (b), show a plot of the 10-step ahead prediction on Adjusted Closing Price of SP500. On this plot also plot out the 95% prediction intervals along with the predicted adjusted closing prices.

```
# predict the future 10 days
forecasts = predict(fit,10)
plot(c(1:10),exp(forecasts$pred),type='b',lty=1,xlab='time ahead',
     ylab='Predicted',xlim=c(0,10),ylim=c(2000,3000),lwd=2,col='red')
lines(c(1:10),exp(forecasts$pred+1.96*forecasts$se),lty=2,lwd=2,col='blue')
lines(c(1:10),exp(forecasts$pred-1.96*forecasts$se),lty=2,lwd=2,col='blue')
legend('topleft', c("Predicted","+/- 1.96 standard errors"), lty=c(1,2),
lwd=c(2,2),col=c("red","blue"), bg="gray90")
```




```
# combine with past 1 year performance
x_axis = seq(from=2019,by=1/252,length=10)
plot(exp(log_price_ts),xlim=c(2018, 2019.5),ylim=c(1000,3000))
lines(x_axis,exp(forecasts$pred),col="red")
lines(x_axis,exp(forecasts$pred+1.96*forecasts$se),lty=2,lwd=2,col='blue')
lines(x_axis,exp(forecasts$pred-1.96*forecasts$se),lty=2,lwd=2,col='blue')
```



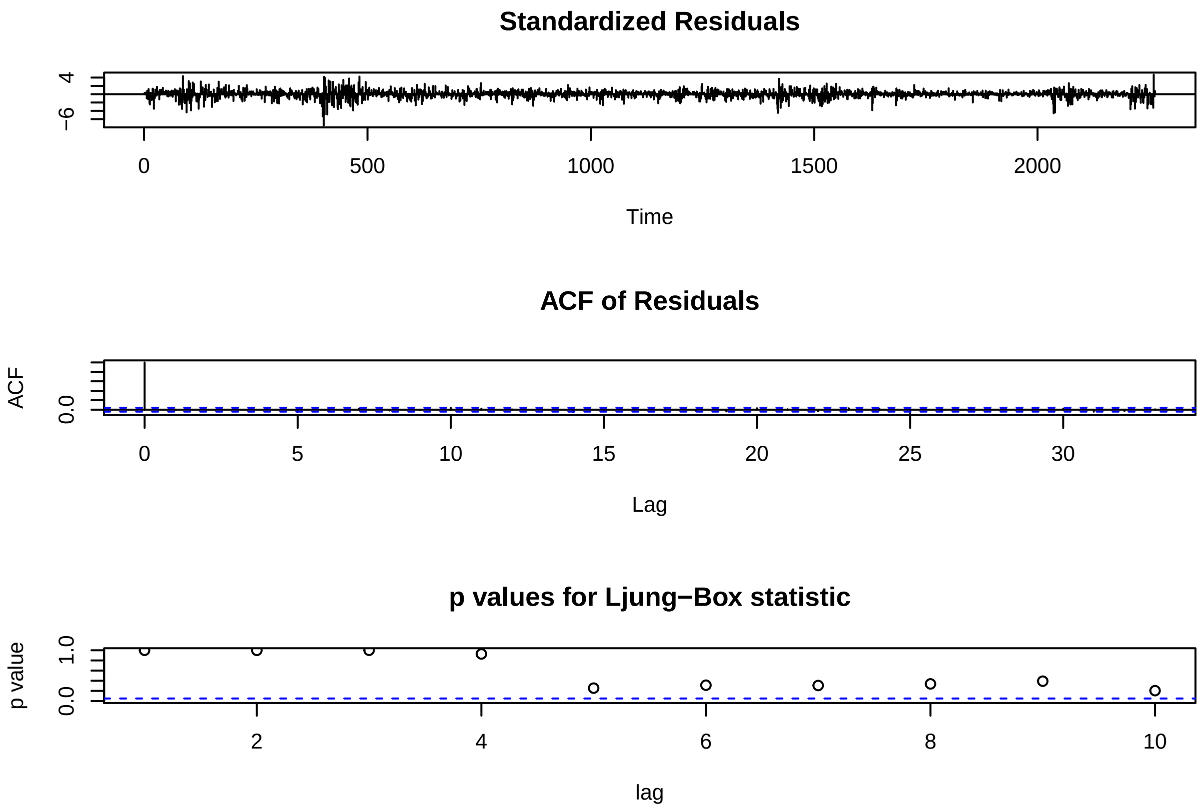

(d) Derive the VaR for $\alpha = .005$ for 1 day ahead relative of 1 million dollar portfolio, based on your analysis in (b) and assuming a portfolio of 1 million dollars. Utilize the normal distribution on the white noise for this derivation, and comment on your thoughts on the accuracy of utilizing such a model based on diagnostics from part (b).

We find the optial arima is (1,0,3), and 1 day ahead VaR is 23648.98. From (b), we mentioned that there is some autocorrelations for lag residuals, and the residuals have heavy tail than normal, meaning the estimatedVaR might be is greater than the actual possible value.

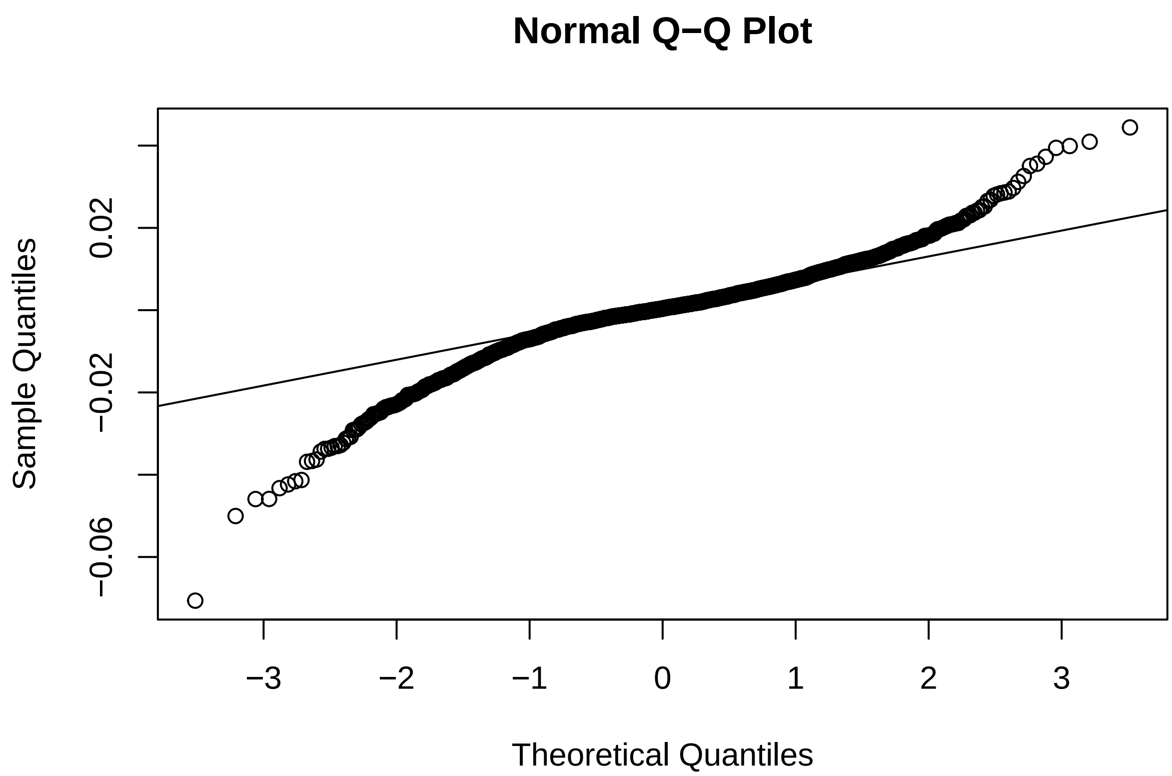
```r
# find new optiaml arima model for log return
auto.arima(log_return,max.P=5,max.Q=5,ic="aic")
```

```
## Series: log_return
## ARIMA(1,0,3) with non-zero mean
##
## Coefficients:
##          ar1      ma1     ma2      ma3   mean
##       0.7820  -0.8286  0.0438  -0.0486  4e-04
## s.e.  0.1388   0.1397  0.0281   0.0228  2e-04
##
## sigma^2 estimated as 8.91e-05:  log likelihood=7343.49
## AIC=-14674.98   AICc=-14674.95   BIC=-14640.64
```

```r
fit_return = arima(log_return,order=c(1,0,3))
tsdiag(fit_return)
```




```r
qqnorm(fit_return$resid)
qqline(fit_return$resid)
```




```r
forecast_return = predict(fit_return,10)
VaR = -(exp(forecast_return$pred[1] + qnorm(.005)*forecast_return$se[1])-1)*1000000
VaR
```

```
## [1] 23648.98
```

## Q3

(a) Carry out an AR(2) model estimation on the square of the log-returns of the SP500 data utilized in Problem 2 to derive estimates for the parameters in an ARCH(2) model - report out your results on these estimated parameters and their interpretation.

From the notes, we can easily deduce that in ARCH(2) model for Xˆ2,$\sigma^2 = \alpha_0 + \alpha_1 * x_{n-1}^2 + \alpha_2 * x_{n-2}^2$, $x_n^2 = (\alpha_0 + \alpha_1 * x_{n-1}^2 + \alpha_2 * x_{n-2}^2)\epsilon_n^2 = (\alpha_0 + \alpha_1 * x_{n-1}^2 + \alpha_2 * x_{n-2}^2)(1 + \epsilon_n^2 - 1) = (\alpha_0 + \alpha_1 * x_{n-1}^2 + \alpha_2 * x_{n-2}^2) + V_n$, where $V_n = \sigma_n^2(\epsilon_n^2 - 1)$. It follows the same form as AR(2).

Based on AR(2) model, the estimated coefficients $\alpha_0$, $alpha_1$, $\alpha_2$ are 1e-04, 0.1642, and 0.3307 respectiely.

```r
ar_model =  arima(log_return^2, order = c(2,0,0), method = "ML")
ar_model
```

```
##
## Call:
## arima(x = log_return^2, order = c(2, 0, 0), method = "ML")
##
## Coefficients:
##          ar1     ar2  intercept
##       0.1642  0.3307      1e-04
## s.e.  0.0198  0.0198      0e+00
##
## sigma^2 estimated as 4.38e-08:  log likelihood = 15960.43,  aic = -31912.86
```

(b) Based on the ARCH(2) model derived in (a) derive the VaR for $\alpha = .005$ for 1 day ahead. Be explicit about the model you are utilizing in computing this value and assume a portfolio of 1 million dollars.

From (a), we can have $\sigma^2_{n+1}$,and then $x^2_{n+1} = \sigma^2_{n+1}\epsilon^2_n$. It is normal distirbution with N(0, $\sigma^2_{n+1}$). So the VaR for alpha = 005 is 25667.57, and we can notice that it is slightly higher than the VaR from last arima (1,0,3) model.

```
# cal sigma^2 for future 1 day
n=length(log_return)
simga_2 = t(c((log_return^2)[n], (log_return^2)[n-1],1)) %*% (ar_model$coef)
VaR2 = -(exp(qnorm(0.005)*sqrt(simga_2))-1)*1000000
VaR2
```

```
##           [,1]
## [1,] 25667.57
```