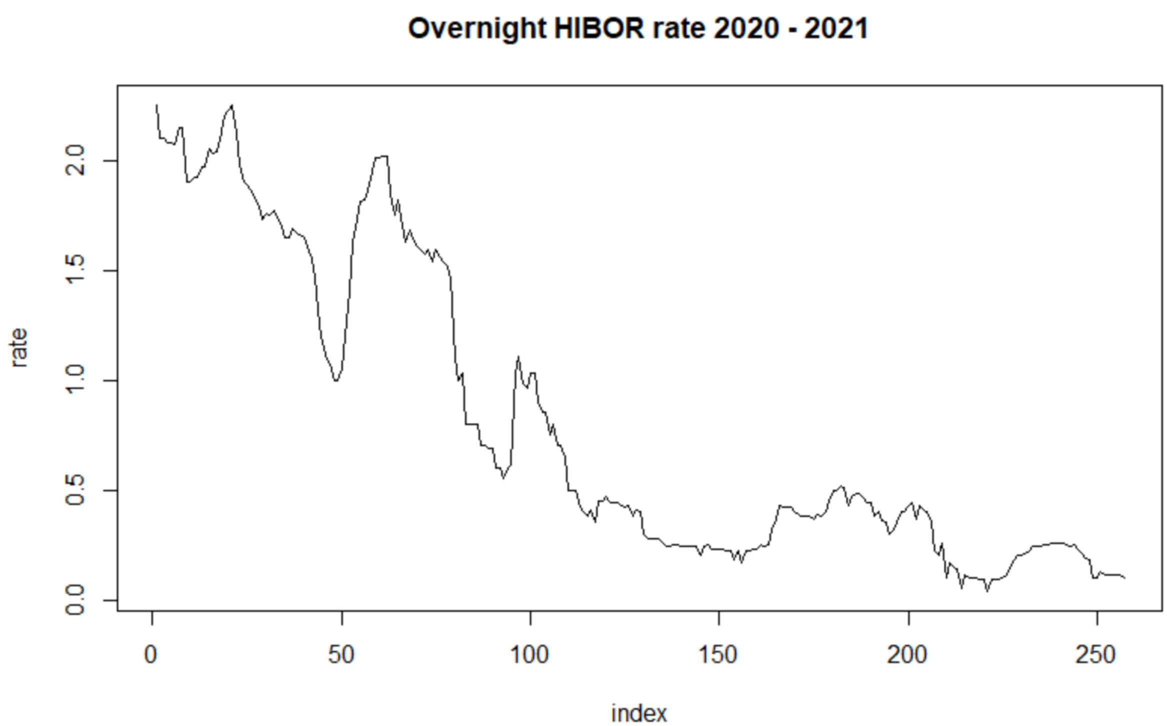
Assignment 6

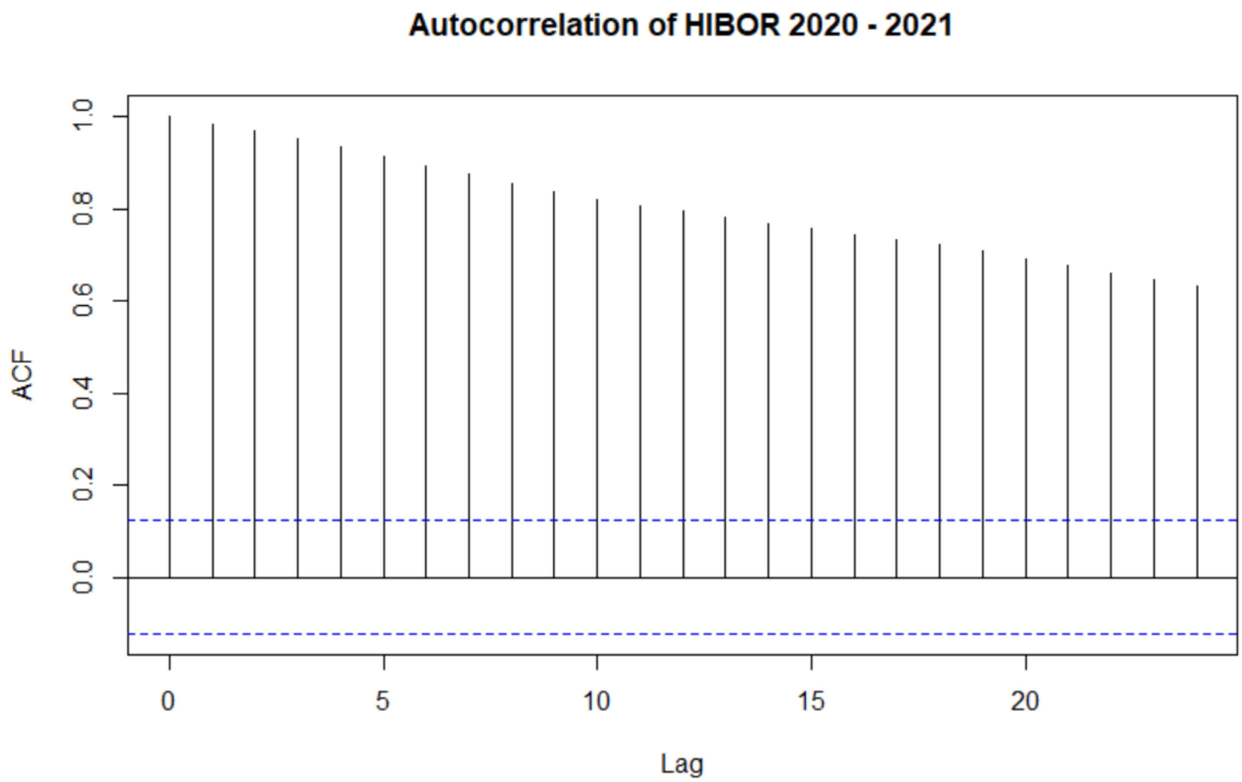
1) Time Series and ACF Plot

```
# 1. Time Series Plot and ACF Plot

par(mfrow=c(1,1))
rate=as.numeric(as.character(HIBOR$rate))
ts.plot(rate,type = "l",xlab = "index",ylab="rate",main = "Overnight HIBOR rate 2020 - 2021")

acf(rate,main="Autocorrelation of HIBOR 2020 - 2021")
```

Time Series



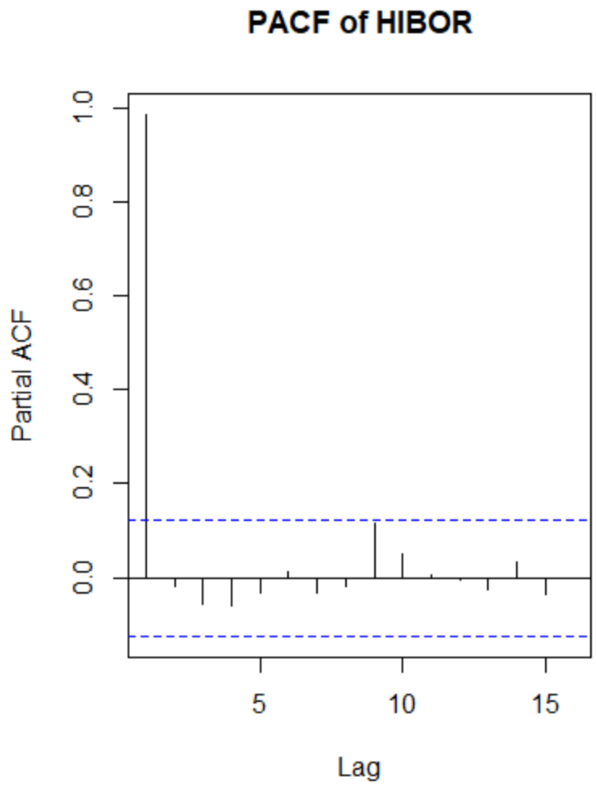
ACF

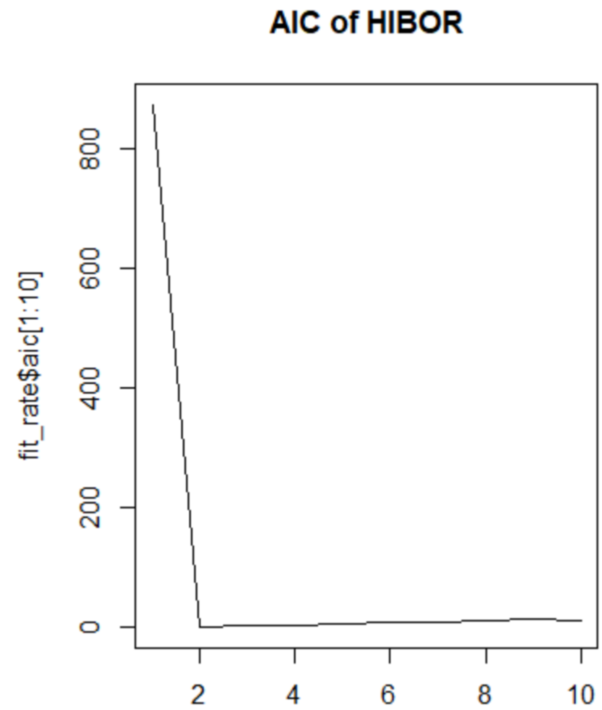


The time series of HIBOR rate seems to be trending and drifting lower from 2020 - 2021. There are also fluctuations that cause the trend to not be clear. As for the autocorrelation function of HIBOR rate, it is decaying very slowly and remains well above the significant zone, which indicates that the time series is non-stationary.

1.2) AR of Training Data

```
# 2. AR Order of Training Data

  # training (index: 1 - 252)
rate_training <- rate[1:252]
  # testing (index: 253 - 257)
rate_testing <- rate[253:257]

  # determine order
par(mfrow=c(1,2))
pacf(rate_training,main="PACF of HIBOR", xlim = c(1,16))
fit_rate <- ar(rate_training)
ts.plot(fit_rate$aic[1:10],xlab=" ",main="AIC of HIBOR")
fit_rate$order
```





The order as determined by PACF and AIC of HIBOR is 1. AR(1) will be used to model the data.

1.3) Fitting AR(1) to Data and Vasicek Model

```
# 3. AR(1) fitted to training data, Vasicek Model

fit_rate <- arima(rate_training,order = c(1,0,0))
fit_rate
```

```
Coefficients:
         ar1  intercept
      0.9975     1.0718
s.e.  0.0031     0.8352

sigma^2 estimated as 0.004599:  log likelihood = 317.9,  aic = -629.8
```

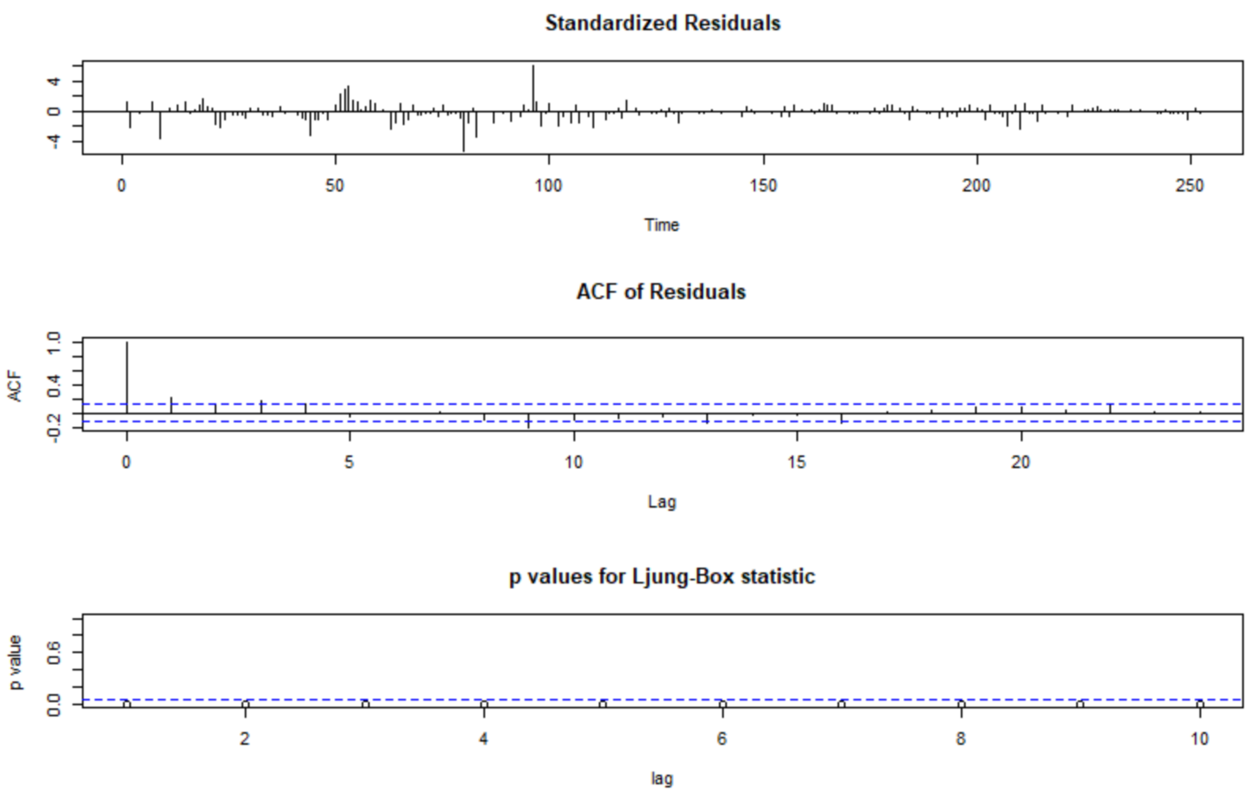
<u>Vasicek Model</u>

$$(r_t - \mu) = (1 - \alpha)(r_{t-1} - \mu) + W_t$$

Parameter	*Estimated*
long-term mean rate level ($\hat{\mu}$)	**1.0718**
speed of mean reversion ($\hat{\alpha}$)	1 - 0.9975 = **0.0025**
sd of noise ($\hat{\sigma}$)	sqrt(0.004599) = **0.0678**
$(r_t - 1.0718) = 0.0025(r_t\text{-}1 - 1.0718) + W_t$	

1.4) Model Diagnostics

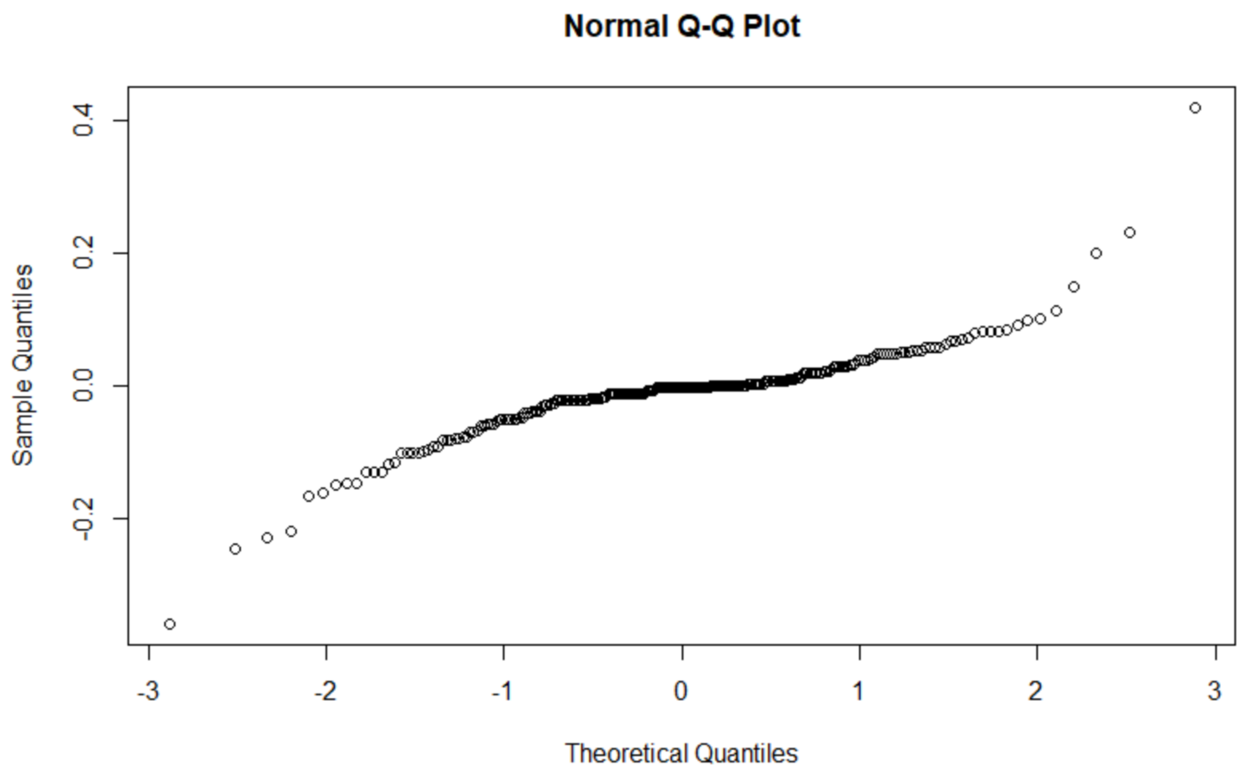
```
# 4. Model Diagnostics

par(mfrow=c(1,1))
tsdiag(fit_rate)
qqnorm(fit_rate$residuals)
```



Standardized Residuals: there are outliers for residuals because some residuals are outside the 3 standard deviation, which indicates heavy tailedness.

ACF of Residuals: at certain lag intervals, the autocorrelation of residuals is marginally significant, which means there may be correlation. Though it seems that the residuals are roughly correlated

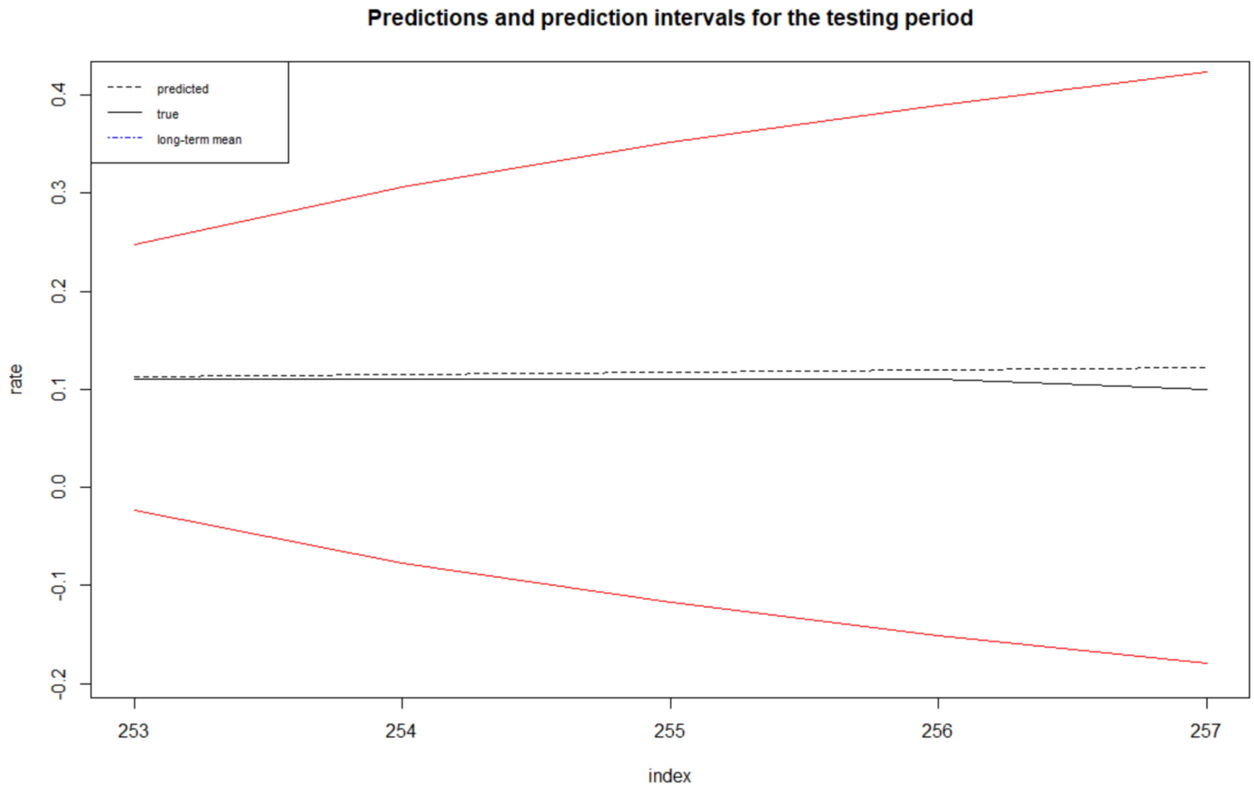
p values for Ljung-Box statistics: none of the p values are significant. The null hypothesis (H_0) at lag 10 is p1 = p2 = p3 ... = p10 = 0 is rejected because the p value is not significant, which indicates that there is correlation among the residuals since correlation is not equal to 0.



Normal Q-Q Plot: there is a normality issue because it does not plot a straight line. Also, there seems to be some outliers at the end points of the graph.

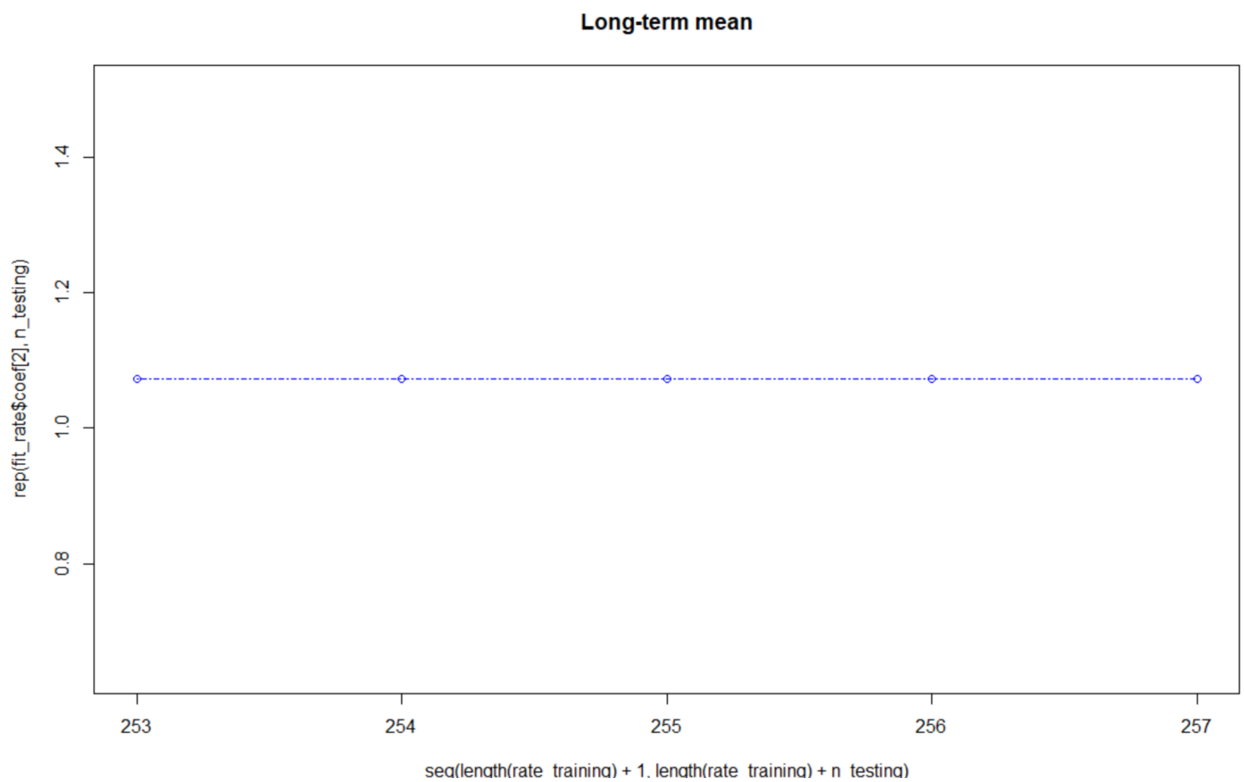
1.5) Predictions and Prediction Interval

```
# 5. Predictions / Prediction Intervals, Comparison to Test Data

  # prediction for HIBOR
n_testing <- length(rate_testing)
predict_rate <- predict(fit_rate, n_testing)

ts.plot(predict_rate$pred,lty = 2, ylim = c(-.19,0.41)) #
lines(seq(length(rate_training) + 1,length(rate_training) + n_testing),rate_testing)

  # prediction interval for HIBOR
PI_l <- predict_rate$pred - 2*predict_rate$se
PI_r <- predict_rate$pred + 2*predict_rate$se
ts.plot(predict_rate$pred,xlab="index",ylab="rate", ylim = c(-.19,0.41), lty = 2,main="Predictions and prediction intervals for the testing period")
lines(PI_r,col = 'red')
lines(PI_l,col = 'red')
lines(seq(length(rate_training) + 1,length(rate_training) + n_testing),rate_testing)
lines(seq(length(rate_training) + 1,length(rate_training) + n_testing),rep(fit_rate$coef[2],n_testing),col="blue",lty=4)
legend('topleft',legend = c('predicted','true','long-term mean'),lty = c(2,1,4), cex = 0.7, col = c('black','black','blue'))
```

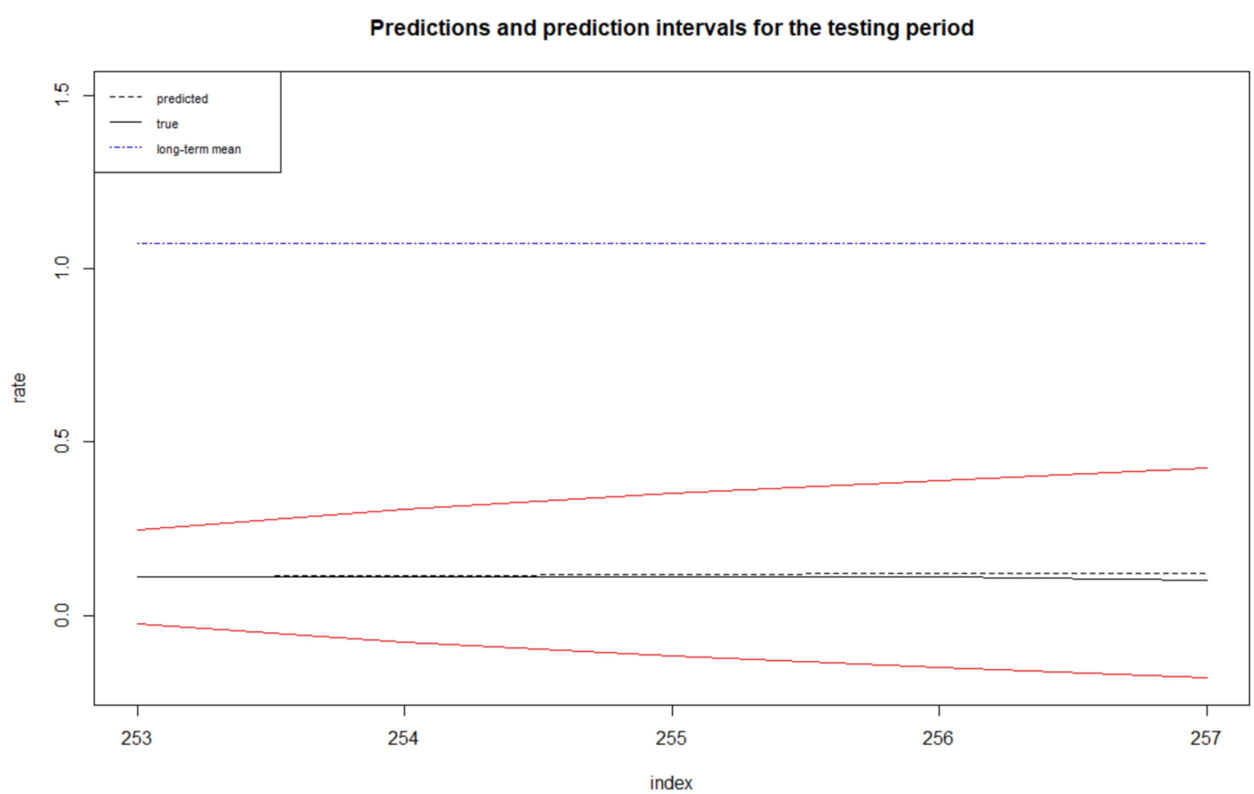


Because the long-term mean determined by the training period is **1.0718** (from part 4 - Model Diagnostics), it is not contained in the interval above. Since the time series is non-stationary, it is expected that the predictions are not centered around the long-term mean.

```
  # long-term mean for HIBOR
plot(seq(length(rate_training) + 1,length(rate_training) + n_testing),rep(fit_rate$coef[2],n_testing),col="blue",lty=4, main="Long-term mean")
lines(seq(length(rate_training) + 1,length(rate_training) + n_testing),rep(fit_rate$coef[2],n_testing),col="blue",lty=4)
```



Tweaking the scale will display the long-term mean outside the prediction intervals.



2) Problem 6.14 on page 418 of the book by Carmona

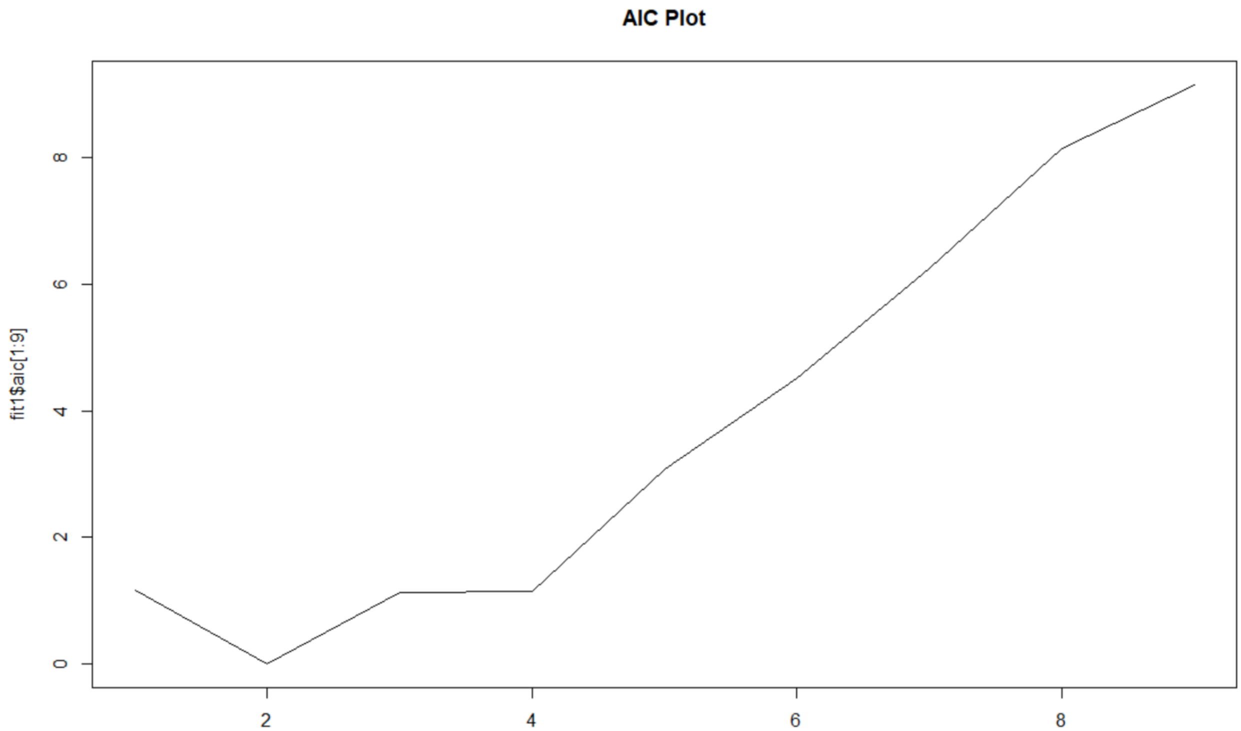
2.1) White Noise and AR(3)

```
# 1. White noise and AR(3) Simulation

  # white noise
set.seed(14)
W<-rnorm(1024)

  # AR(3)
X<-arima.sim(1024,model=list(ar=0.07,0.02,0.3),n.start=1,start.innov=c(0), innov = W)
```

2.2) AR fitting and applications

```
# 2. AR fitting and applications

  # AR order 9 w/ AIC
par(mfrow=c(1,1))
fit1<-ar(X)
ts.plot(fit1$aic[1:9],xlab=" ",main="AIC Plot")

  # fit AR(1)
fit2<-arima(X,order=c(1,0,0))
fit2
tsdiag(fit2)

  # forecasting T+16
pred <- predict(fit2, n.ahead = 16)
PI_l <- pred$pred - 2*pred$se
PI_r <- pred$pred + 2*pred$se
ts.plot(pred$pred, ylim = c(-2.5,2.6))
lines(PI_r,col = 'red')
lines(PI_l,col = 'red')

pred$pred
```

AIC



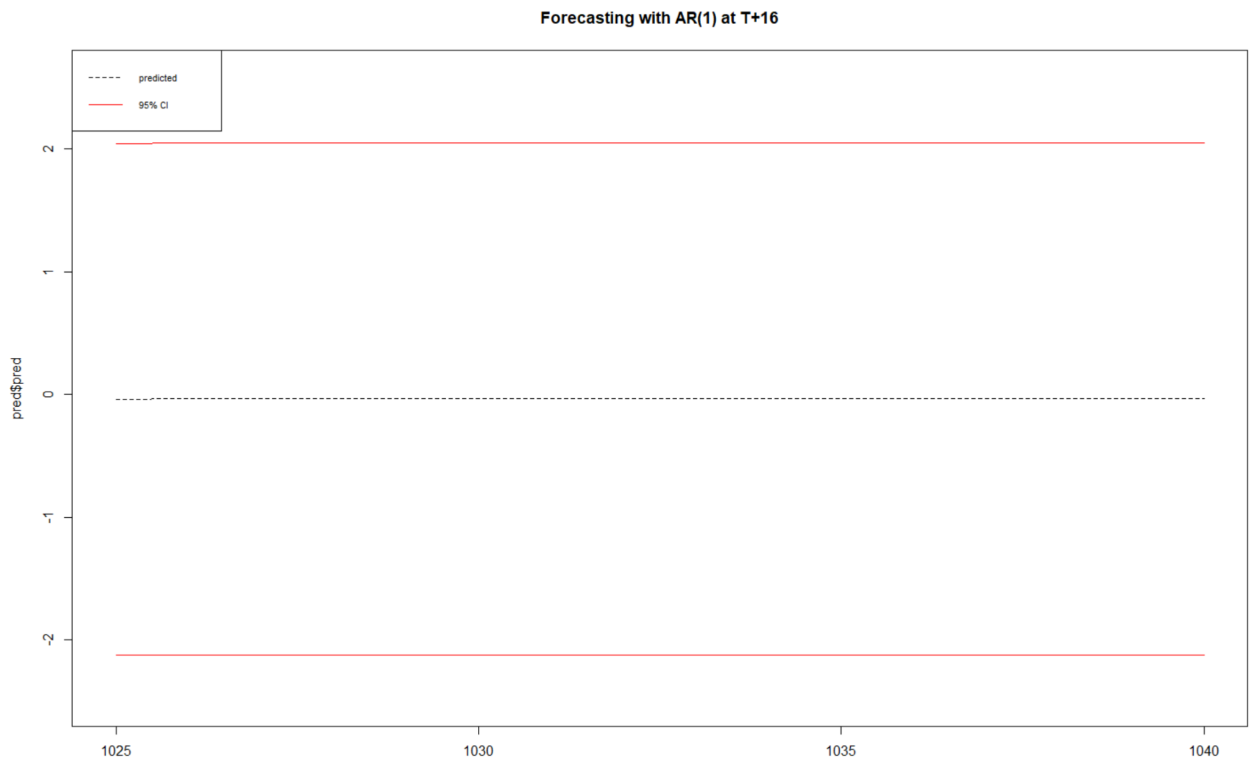
Based on the fitted AR models up to order 9, according to AIC, the best model with the least AIC is at **order 1**.

Model Coefficients

```
Coefficients:
         ar1  intercept
      0.0555    -0.0355
s.e.  0.0312     0.0345

sigma^2 estimated as 1.087:  log likelihood = -1495.58,  aic = 2997.16
```

AR(1) = $(X_t + 0.0355) = 0.0555(X_{t-1} + 0.00355) + W_t$

Forecasting




```
> pred$pred
Time Series:
Start = 1025
End = 1040
Frequency = 1
 [1] -0.04017850 -0.03571422 -0.03546661 -0.03545288 -0.03545212 -0.03545208 -0.03545208 -0.03545208 -0.03545208 -0.03545208 -0.03545208 -0.03545208
[13] -0.03545208 -0.03545208 -0.03545208 -0.03545208
```

Prediction values around the long-term mean of -0.0355.

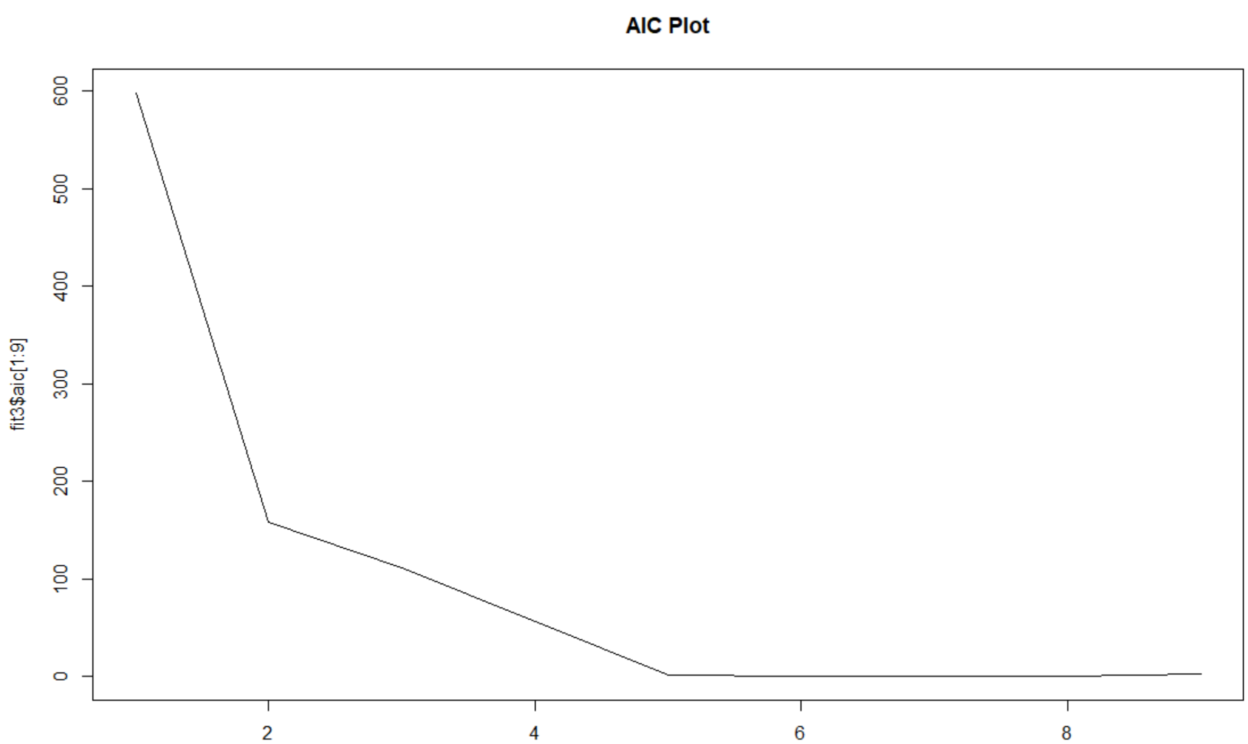
2.3) White Noise and ARMA(3,4)

```
# 3. White noise and ARMA(3,4) Simulation

  # white noise
set.seed(14)
W<-rnorm(1024)

  # ARMA(3,4)
X2 <- arima.sim(n = 1000, list(ar = c(0.07,0.02,0.3), ma = c(0.4,0.3,0.2,0.05)), innov = W)
ts.plot(X2)
```

2.4) AR Fitting and Applications (2)

```r
# 4. AR fitting and applications

  # AR order 9 w/ AIC
fit3<-ar(X2)
ts.plot(fit3$aic[1:9],xlab=" ",main="AIC Plot")

  # fit AR(4)
fit4<-arima(X2,order=c(4,0,0))
fit4
tsdiag(fit2)

  # forecasting T+16
pred <- predict(fit4, n.ahead = 16)
PI_l <- pred$pred - 2*pred$se
PI_r <- pred$pred + 2*pred$se
ts.plot(pred$pred, ylim = c(-3,3), lty = 2, main="Forecasting with AR(4) at T+16")
lines(PI_r,col = 'red')
lines(PI_l,col = 'red')
legend('topleft',legend = c('predicted','95% CI'),lty = c(2,1), cex = 0.7, col = c('black', 'red'))

pred$pred
```

<u>AIC</u>



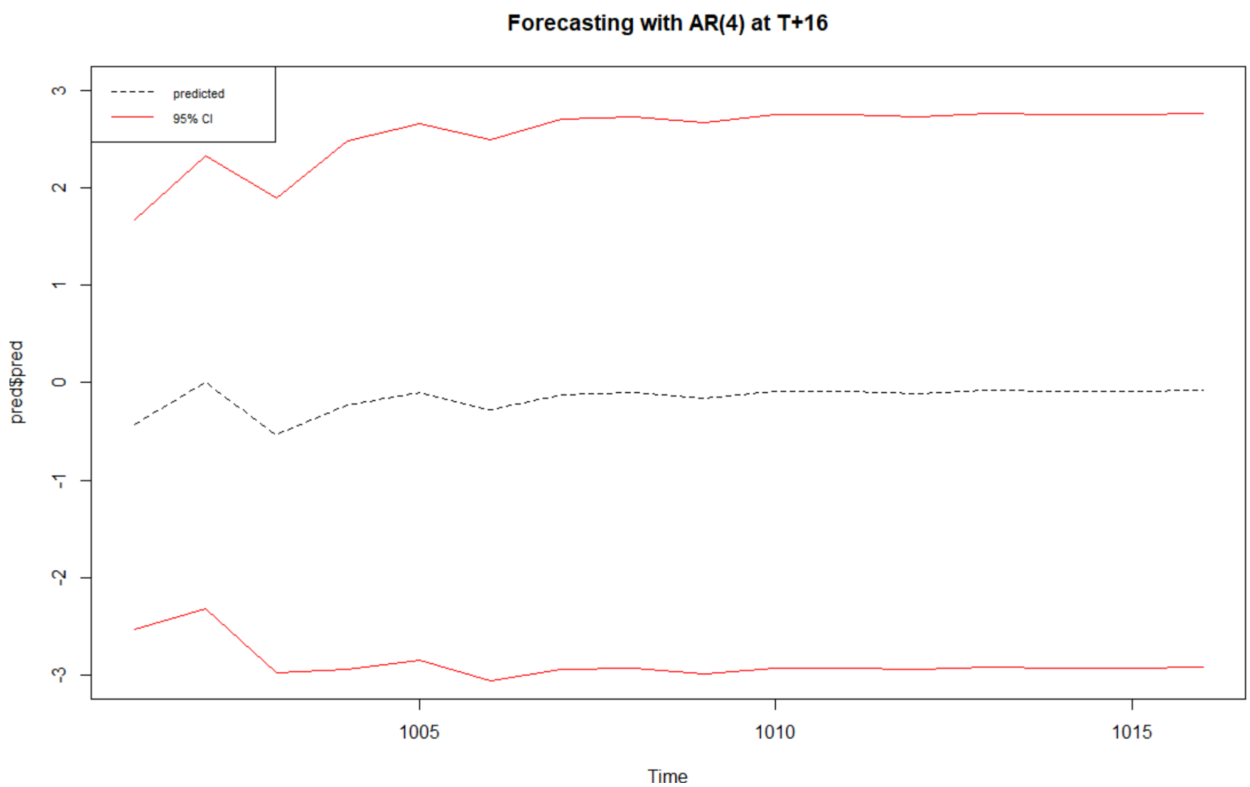
Based on the fitted AR models up to order 9, according to AIC, the best model with the least AIC while maintaining model complexity is at **order 4**.

<u>Model Coefficients</u>

```
Coefficients:
         ar1     ar2     ar3      ar4  intercept
      0.4716  0.1291  0.3370  -0.2335    -0.0812
s.e.  0.0308  0.0325  0.0325   0.0309     0.1118

sigma^2 estimated as 1.1:  log likelihood = -1467.08,  aic = 2946.16
```

AR(4) = $(X_t + 0.0812) = 0.4716(X_{t-1} + 0.0812) + 0.1291(X_{t-2} + 0.0812) + 0.3370(X_{t-3} + 0.0812) - 0.2335(X_{t-3} + 0.0812) + W_t$

<u>Forecasting</u>



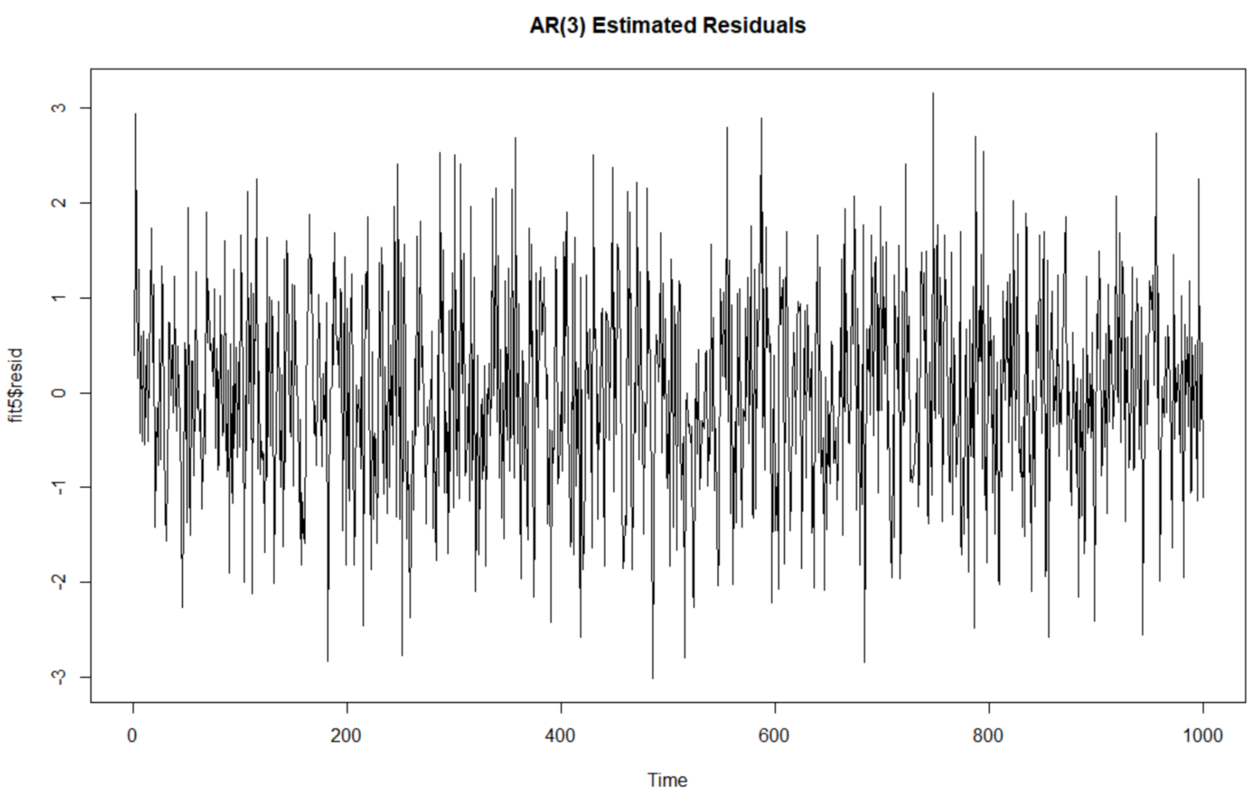

```
> pred$pred
Time Series:
Start = 1001
End = 1016
Frequency = 1
 [1] -0.431707814  0.003842957 -0.536923286 -0.230216531 -0.099794823 -0.282619972 -0.122389948 -0.098091967 -0.158010687 -0.086451365
[11] -0.089664013 -0.107806816 -0.078673923 -0.085068201 -0.089686828 -0.078637296
```

Prediction values around the long-term mean of -0.0812.

2.5) ARMA Fitting and Applications

```r
# 5. ARMA Fitting and applications

  # AR(3) forced fitting
fit5 <- arima(X2,order=c(3,0,0))

  # estimated residuals
plot(fit5$resid, main="AR(3) Estimated Residuals")
acf(fit5$resid,main="ACF of AR(3) Residuals")

  # MA Fit
MAfit1<-arima(fit5$resid,order=c(0,0,1))
MAfit2<-arima(fit5$resid,order=c(0,0,2))
MAfit3<-arima(fit5$resid,order=c(0,0,3))
MAfit4<-arima(fit5$resid,order=c(0,0,4))
MAfit5<-arima(fit5$resid,order=c(0,0,5))
round(c(MAfit1$aic,MAfit2$aic,MAfit3$aic,MAfit4$aic,MAfit5$aic),3)

  # forecasting T+16

fit6 <- arima(X2,order=c(3,0,5))
fit6
pred <- predict(fit6, n.ahead = 16)
PI_l <- pred$pred - 2*pred$se
PI_r <- pred$pred + 2*pred$se
ts.plot(pred$pred, ylim = c(-3,3), lty = 2, main="Forecasting with ARMA(3,5) at T+16")
lines(PI_r,col = 'red')
lines(PI_l,col = 'red')
legend('topleft',legend = c('predicted','95% CI'),lty = c(2,1), cex = 0.7, col = c('black', 'red'))

pred$pred
```

Estimated Residuals



Moving Average Fitting

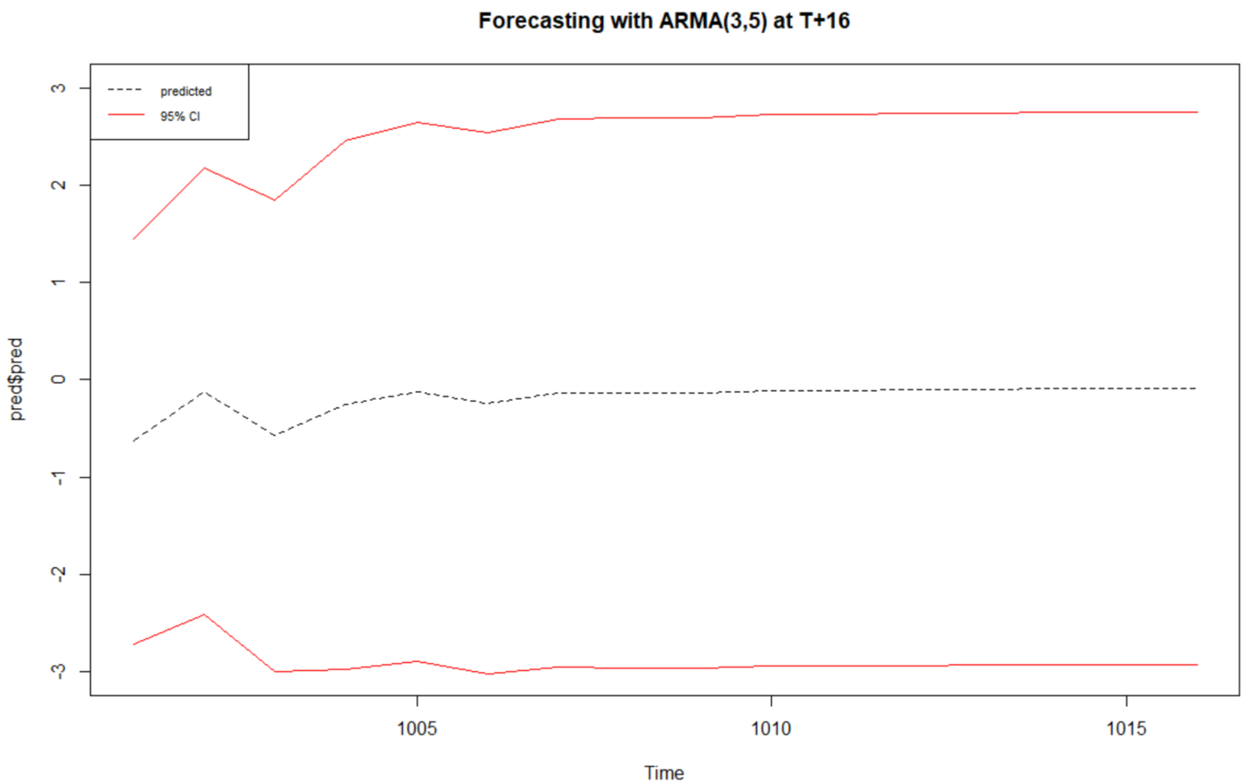
```
> round(c(MAfit1$aic,MAfit2$aic,MAfit3$aic,MAfit4$aic,MAfit5$aic),3)
[1] 2991.961 2990.124 2963.136 2952.890 2939.733
```

The moving average with the lowest AIC is MA(5) at 2939.733. Note that MA(4) has an AIC of 2952.89, which is also a relatively low AIC. MA(4) and MA(5) have an AIC difference of ~13, but for simplicity, the model with the lowest AIC will be chosen, which is **MA(5)**.

Model Coefficients

```
Coefficients:
          ar1     ar2     ar3     ma1     ma2     ma3     ma4      ma5  intercept
      -0.0022  0.2865  0.2307  0.4618  0.0911  0.2288  0.0407  -0.1302    -0.0808
s.e.   0.2075  0.1809  0.0933  0.2076  0.1731  0.0651  0.0982   0.0765     0.1147

sigma^2 estimated as 1.087:  log likelihood = -1461.04,  aic = 2942.08
```

Forecasting




```
> pred$pred
Time Series:
Start = 1001
End = 1016
Frequency = 1
 [1] -0.63262702 -0.12170788 -0.57404847 -0.25760466 -0.12164700 -0.24515286 -0.13291458 -0.13717393 -0.13351058 -0.10884111 -0.10882967
[12] -0.10091767 -0.09523986 -0.09298347 -0.08953645 -0.08758771
```

AR(1) and ARMA(3,5) Comparison

Both the models have very similar intercepts, which means that the long-term mean should be around -0.08.

Since both models are not an ARMA(3,4) model the coefficients will not be able to exactly match the true model. The ARMA(3,5) models the time series better than AR(1) because it captures the movement of the residuals under a MA(5) process. Additionally, the time series moves under an AR(3) process which is captured sufficiently by ARMA(3,5). The forecasts for the AR(1) model is constant and does not deviate much, while the forecasts for ARMA(3,5) shifts. **ARMA(3,5)** has better results in modelling and forecasting.