Time Series HW3

Ze YANG

9/25/2018

Ruppert & Matteson

Problem 1

```r
# Load data
library(tseries)
library(forecast)
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'default/
## America/New_York'
```

```r
library(Ecdat)
```

```
## Loading required package: Ecfun
```

```
##
## Attaching package: 'Ecfun'
```

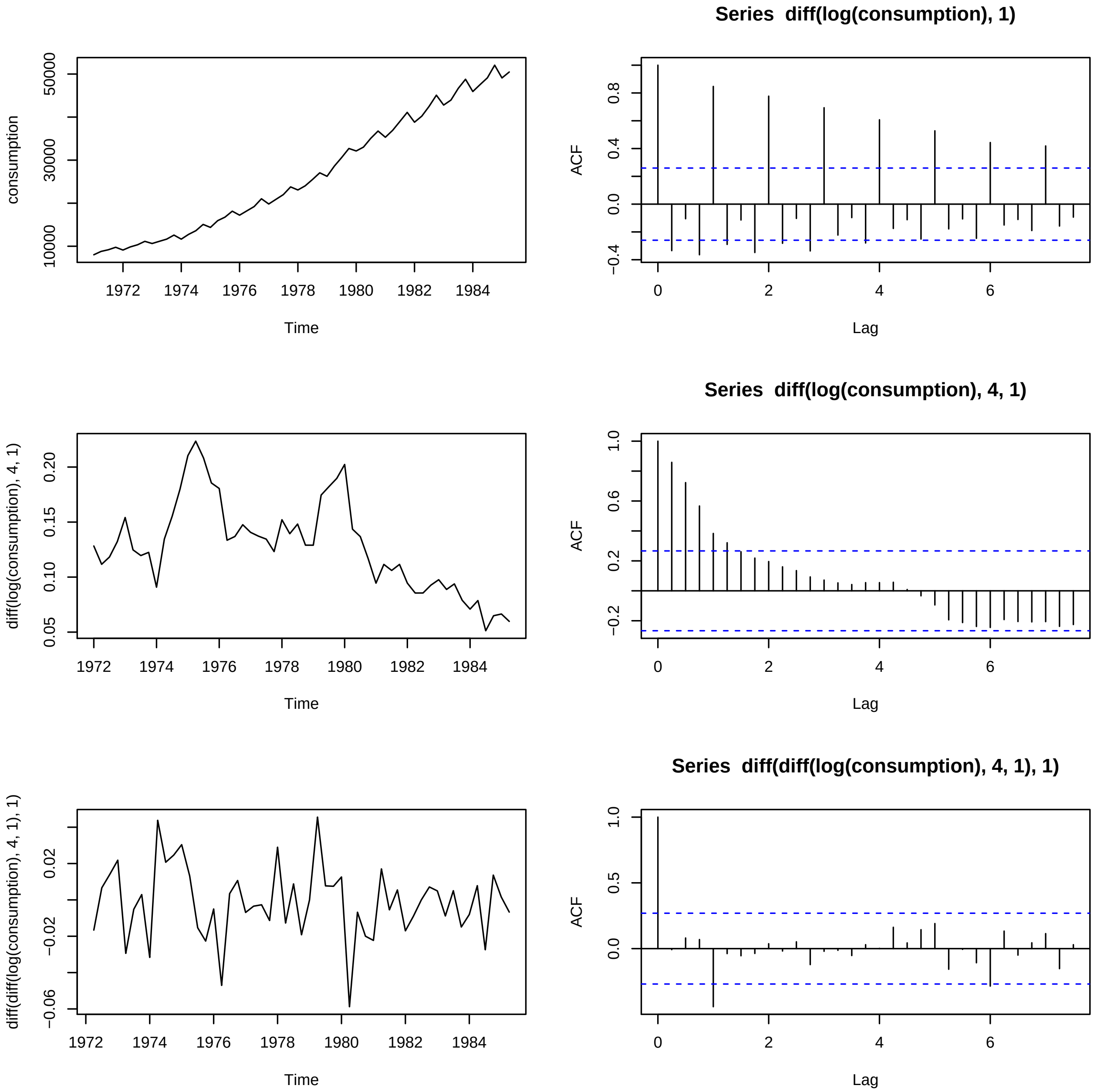
```
## The following object is masked from 'package:forecast':
##
##     BoxCox
```

```
## The following object is masked from 'package:base':
##
##     sign
```

```
##
## Attaching package: 'Ecdat'
```

```
## The following object is masked from 'package:datasets':
##
##     Orange
```

```r
data(IncomeUK)
consumption = IncomeUK[,2]
par(mfrow=c(3,2))
plot(consumption)
acf(diff(log(consumption), 1), 30)
plot(diff(log(consumption), 4, 1))
acf(diff(log(consumption), 4, 1), 30)
plot(diff(diff(log(consumption), 4, 1), 1))
acf(diff(diff(log(consumption), 4, 1), 1), 30)
```



The behavior of consumption

1. The size of seasonal oscillations seem to *increase* with time.
2. There is an upward trend (hence nonstationarity).
3. There is an evident seasonal component. The ACF of the first-difference suggests the seasonality has a period = 4.
4. After applying lag-4 difference, the series still has a trend.
5. After applying the first difference of the lag-4 difference, the series appears to be nice.

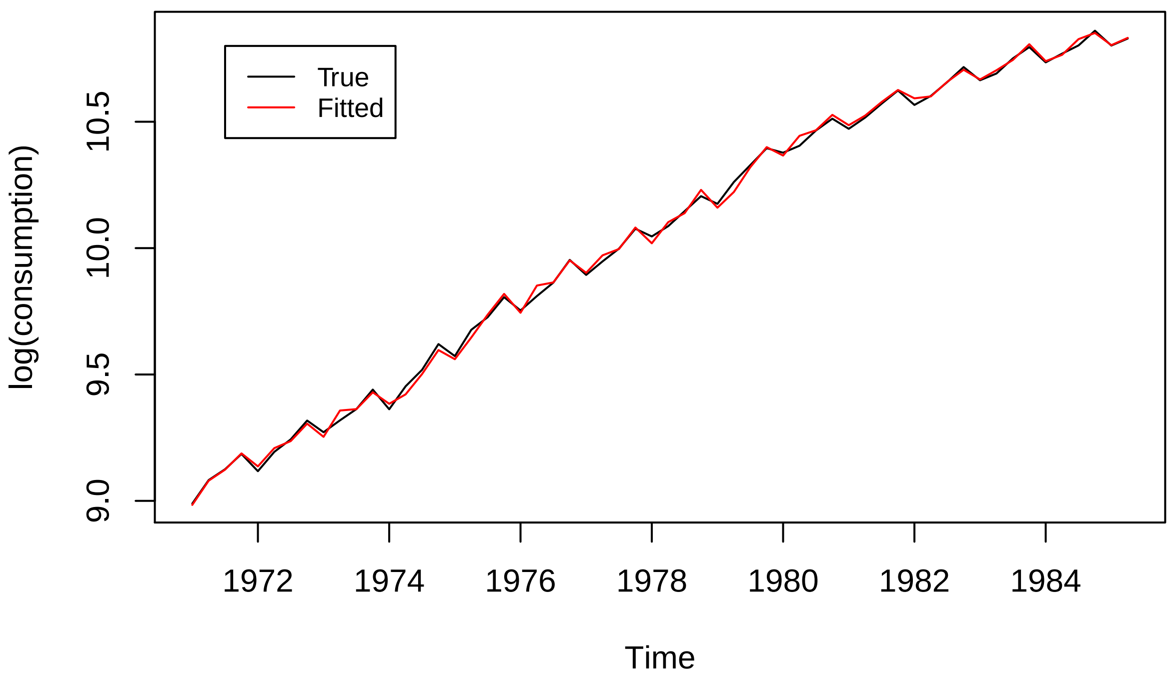
Behavior 1 suggests we apply a Box-Cox (log) transformation to stablize the size of seasonal oscillations. Behavior 2-5 implies we apply both seasonal and non-seasonal differencing to remove two types of serial dependencies.

Problem 2

```r
fit.res = arima(log(consumption),
                order=c(1,1,1),
                seasonal=list(order=c(1,1,0), period=4))
summary(fit.res)
```

```
##
## Call:
## arima(x = log(consumption), order = c(1, 1, 1), seasonal = list(order = c(1,
##     1, 0), period = 4))
##
## Coefficients:
##          ar1      ma1     sar1
##       0.6027  -0.5201  -0.4644
## s.e.  0.4122   0.4264   0.1238
##
## sigma^2 estimated as 0.0003058:  log likelihood = 138.78,  aic = -269.57
##
## Training set error measures:
##                        ME       RMSE        MAE         MPE      MAPE
## Training set -0.001569978 0.01692169 0.01285601 -0.01454162 0.1292932
##                   MASE        ACF1
## Training set 0.2263079 -0.06234695
```

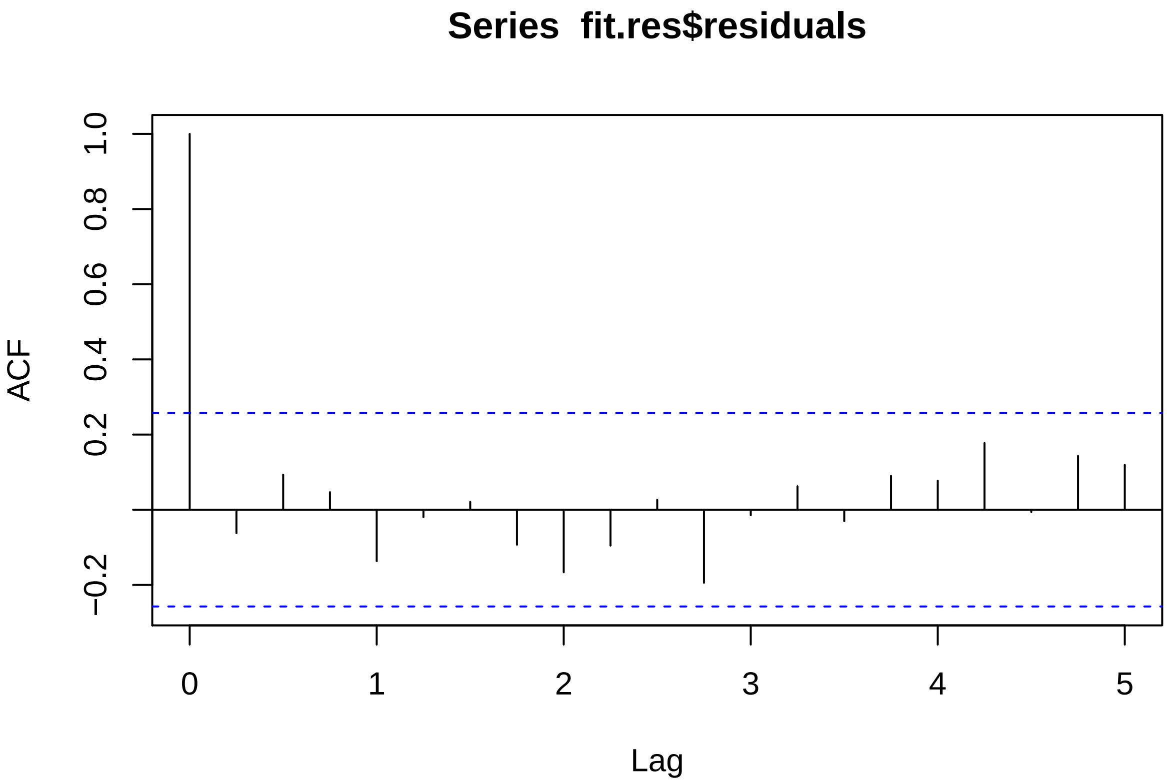
```r
plot(log(consumption))
lines(fitted(fit.res), col='red')
legend(1971.5, 10.8,
       legend=c("True", "Fitted"),
       col=c("black", "red"),
       lty=c(1,1), cex=0.8)
```



An $ARIMA((1,1,1) \times (1,1,0)_4)$ model gives a good fit to the data.

Problem 3

```
acf(fit.res$residuals, 20)
```




```
Box.test(fit.res$residuals, fitdf=3, lag=20, type="Ljung-Box")
```

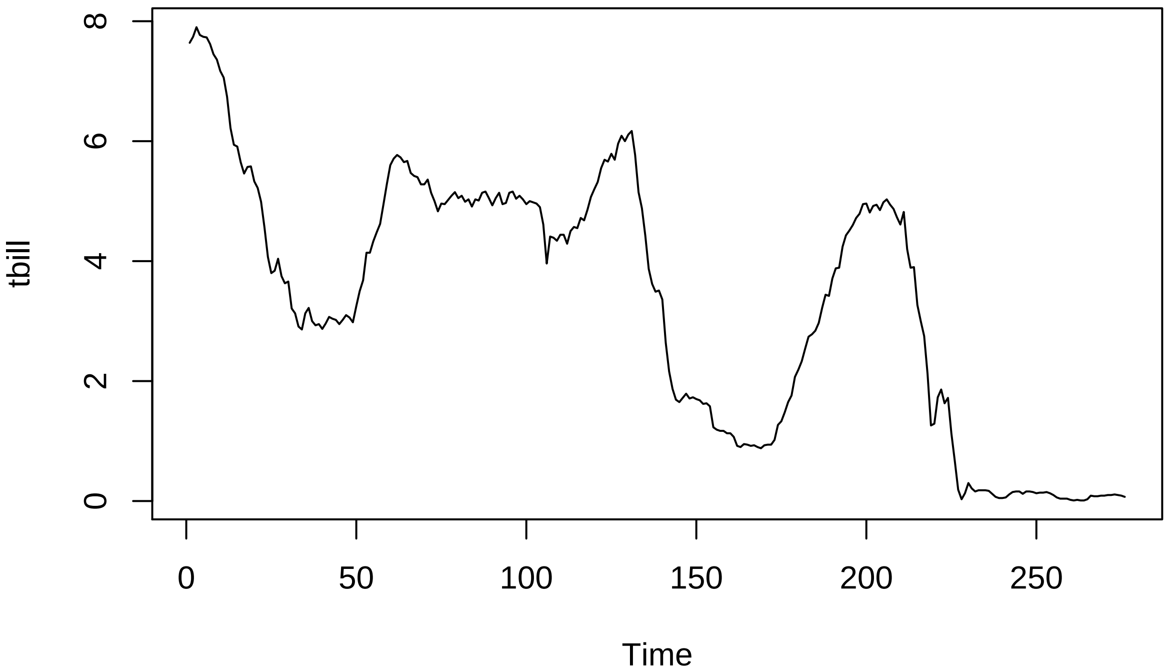
```
##
##  Box-Ljung test
##
## data:  fit.res$residuals
## X-squared = 15.551, df = 17, p-value = 0.5558
```

- The residual ACF displays no evidence of correlation till lag 20. There is no residual autocorrelation detected.
- The p-value of Ljung-Box tests for the residuals is very big. Fail to reject null: no residual autocorrelation.

Three-Month Treasury Bills

(a)

```
tbill.full = read.csv('TB3MS.csv')
tbill = tbill.full[(as.Date(tbill.full$DATE) <= '2012-12-31') & (
  as.Date(tbill.full$DATE) >= '1990-1-1'),]
tbill = ts(tbill$TB3MS)
plot(tbill)
```

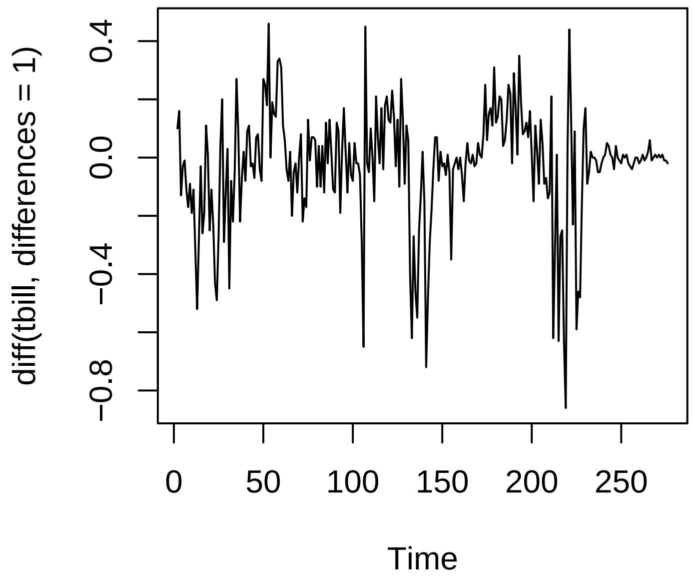


- A clear non-mean-reverting trend can be observed. The orginal series is not stationary.

(b)

```
par(mfrow=c(1,2))
plot(diff(tbill, differences=1))
adf.test(diff(tbill, differences=1))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(tbill, differences = 1)
## Dickey-Fuller = -3.5576, Lag order = 6, p-value = 0.03763
## alternative hypothesis: stationary
```



We choose to use the First difference of the log-transformed series. The ADF test yield 0.03763 p-value, indicates a rejection of the null hypothesis (H_0: the series is integrated) at 95% confidence level. And the plot of the series seems to be approximately stationary indeed.

(c)

```
fit.tbill = auto.arima(
  tbill, d=1, D=1, max.p=10, max.q=10,
  ic='aicc', approx=F, step=F)
summary(fit.tbill)
```

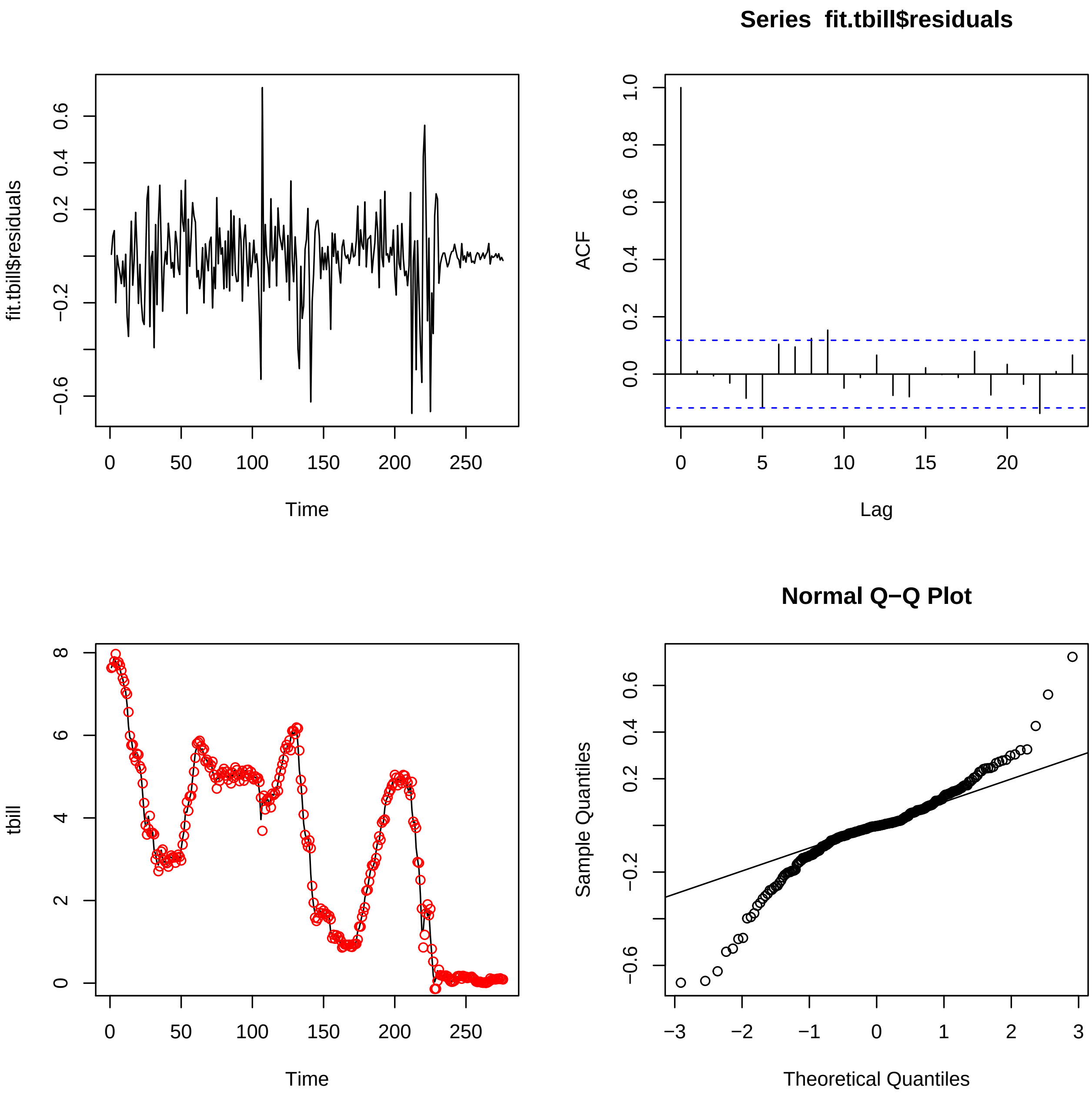
```
## Series: tbill
## ARIMA(3,1,0)
##
## Coefficients:
##          ar1      ar2     ar3
##       0.4093  -0.0300  0.2523
## s.e.  0.0582   0.0633  0.0582
##
## sigma^2 estimated as 0.02894:  log likelihood=98.17
## AIC=-188.34   AICc=-188.19   BIC=-173.87
##
## Training set error measures:
##                       ME      RMSE       MAE      MPE     MAPE     MASE
## Training set -0.01025191 0.1688823 0.1119726 1.736317 11.60244 0.849916
##                    ACF1
## Training set 0.01019362
```

The selected model is $ARIMA(3, 1, 0)$.

(d)

```
par(mfrow=c(2,2))
plot(fit.tbill$residuals)
```

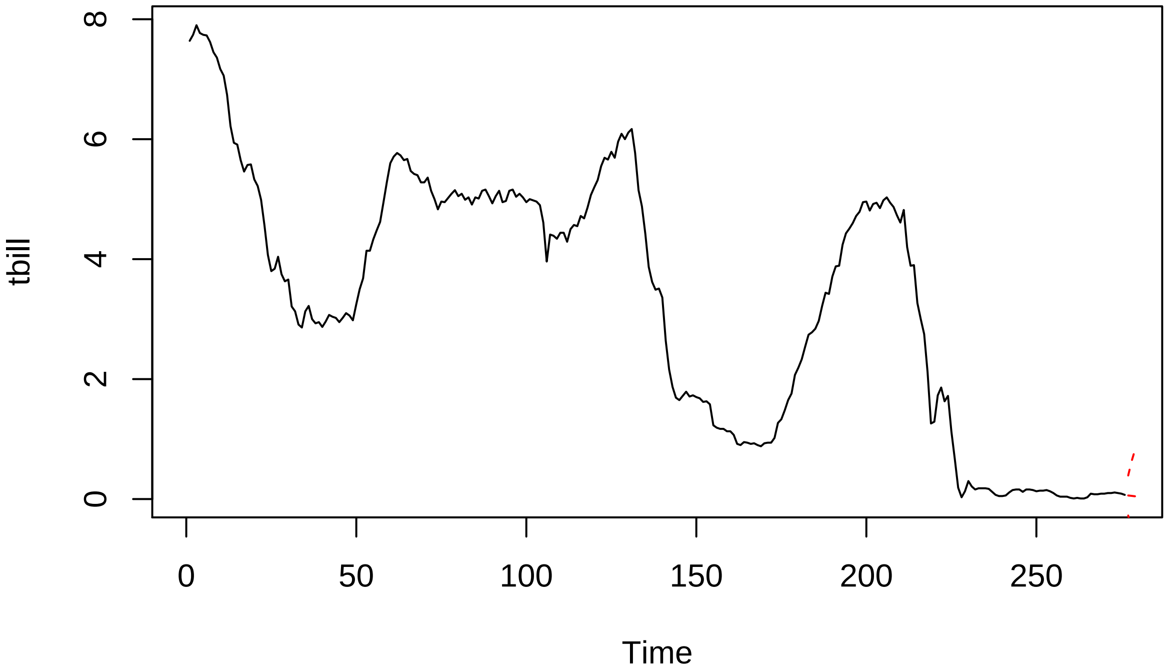
```r
acf(fit.tbill$residuals)
plot(tbill)
points(fitted(fit.tbill), col='red')
legend(0, 15.0,
       legend=c("True", "Fitted"),
       col=c("black", "red"),
       lty=c(1, NA), pch=c(NA, 1), cex=0.8)
qqnorm(fit.tbill$residuals)
qqline(fit.tbill$residuals)
```



The model seems to be a reasonable fit. From upper left to bottom right are: plot of the residuals, ACF of the residuals, true V.S. fitted in-sample, Q-Q plot of the residuals. There is no significant autocorrelation in residuals, and the in-sample fit seems to be good.

(e)

```
pred.res = predict(fit.tbill, n.ahead = 3)
plot(tbill)
lines(pred.res$pred, col='red')
lines(pred.res$pred+1.96*pred.res$se, col='red', lty=2)
lines(pred.res$pred-1.96*pred.res$se, col='red', lty=2)
```



(f)

```
upper = pred.res$pred+1.96*pred.res$se
lower = pred.res$pred-1.96*pred.res$se
for(i in 1:3) {
  print(paste('Center:', pred.res$pred[i],
              '95% CI: [', upper[i], lower[i], ']'))
}
```

```
## [1] "Center: 0.0595900605467794 95% CI: [ 0.393024387527265 -0.273844266433706 ]"
## [1] "Center: 0.0534053210798643 95% CI: [ 0.629598762635297 -0.522788120475569 ]"
## [1] "Center: 0.0461391567090441 95% CI: [ 0.819463838516595 -0.727185525098507 ]"
```

The point forecast and 95% CIs of the subsequent 3 months are printed as above.

```
tbill.test = tbill.full[(as.Date(tbill.full$DATE) > '2012-12-31') & (
  as.Date(tbill.full$DATE) <= '2013-3-30'),]
tbill.test
```

```
##           DATE TB3MS
## 949 2013-01-01  0.07
## 950 2013-02-01  0.10
## 951 2013-03-01  0.09
```

We conclude that all the three out-of-sample true values are contained in the CIs.