FA_Assignment3_Duo_Zhou

Duo Zhou

7/13/2020

This assignment helps understanding linear models for time series

1. Exercise 2 on page 125

Use the file m-dec125910-6111.txt

```
datapath<-"C:/Users/zd000/Desktop/MSCA/Financial Analytics/Assignments/week3/"
dat<-read.table(paste(datapath,"m-dec125910-6111.txt",sep="/"),header=T)
head(dat)
```

```
##       date      dec1      dec2      dec5      dec9     dec10
## 1 19610131  0.058011  0.067392  0.081767  0.096754  0.087207
## 2 19610228  0.029241  0.042784  0.055524  0.056564  0.060245
## 3 19610330  0.025896  0.025474  0.041304  0.060563  0.071875
## 4 19610428  0.005667  0.001365  0.000780  0.011911  0.023328
## 5 19610531  0.019208  0.036852  0.049590  0.046248  0.050362
## 6 19610630 -0.024670 -0.025225 -0.040046 -0.050651 -0.051434
```

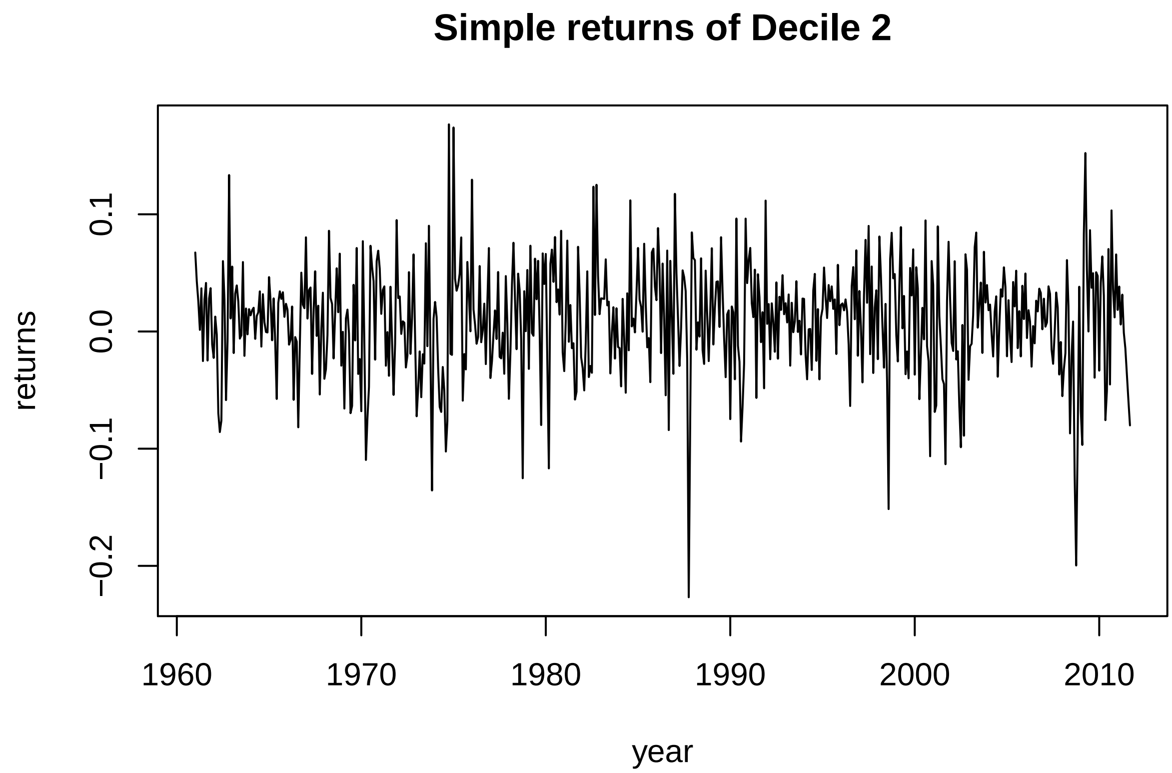
```
dim(dat)
```

```
## [1] 609   6
```

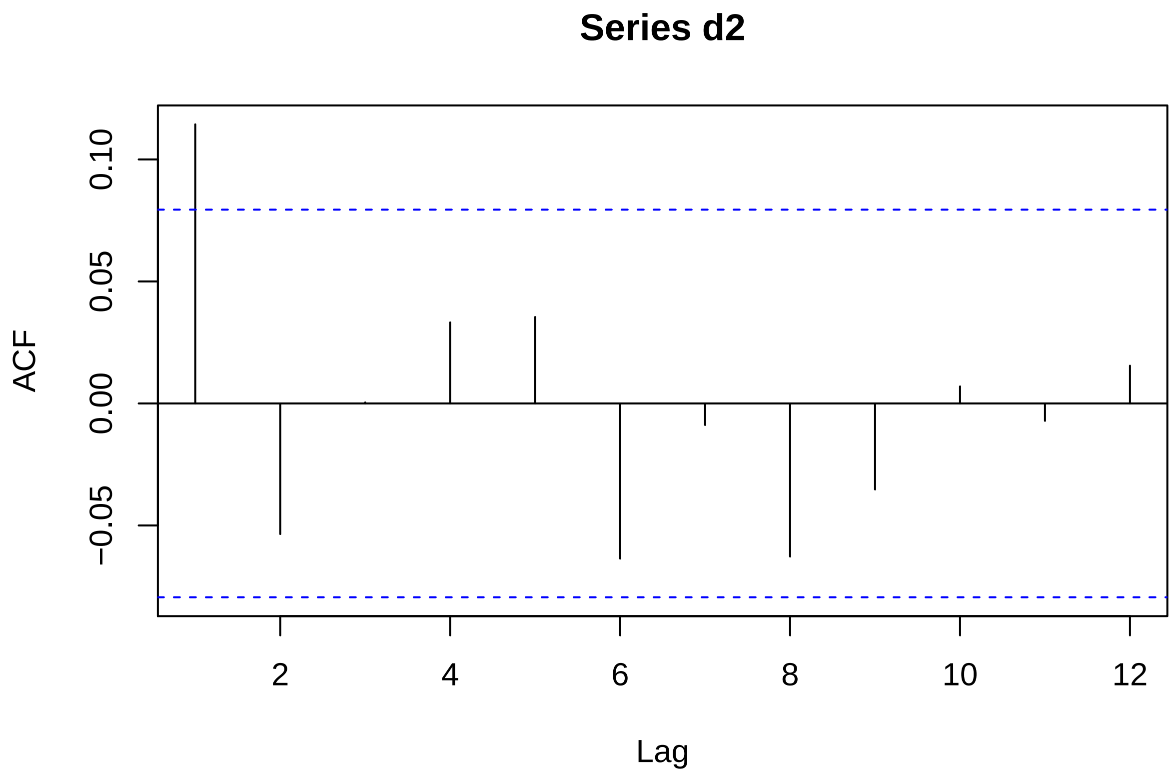
For CRSP portfolios of Decile 2 and Decile 10 returns test null hypothesis that the first 12 lags of autocorrelations equal zero with 5% level

Decile 2

```
# decile 2 simple returns
d2 <- dat$dec2
d2.ts <- ts(d2,frequency = 12, start=c(1961, 1))
# Plot the time series and its ACF for simple returns
plot(d2.ts, xlab="year", ylab="returns",type="l")
title(main = "Simple returns of Decile 2")
```




```
f1=acf(d2,lag=12)
```




```
tt=f1$acf*sqrt(length(d2))
critial.value=rep(qnorm(.975,0,1),length(tt))
compare=cbind(tt,qnorm(.975,0,1),tt>critial.value)
colnames(compare)=c('tt','critical value','tt > critial.value')
compare
```

Z-Score Test

```
##                tt critical value tt > critial.value
##  [1,]  2.82269150       1.959964                  1
##  [2,] -1.32099739       1.959964                  0
##  [3,]  0.01041905       1.959964                  0
##  [4,]  0.81960225       1.959964                  0
##  [5,]  0.87408291       1.959964                  0
##  [6,] -1.56906470       1.959964                  0
##  [7,] -0.21820622       1.959964                  0
##  [8,] -1.54819123       1.959964                  0
##  [9,] -0.86986228       1.959964                  0
## [10,]  0.17229951       1.959964                  0
## [11,] -0.17597701       1.959964                  0
## [12,]  0.38212346       1.959964                  0
```

Based on Z-scores test, only the first lag can reject the null hypothesis that autocorrelation equals zero with 5% level. The rest can not reject the null hypothesis.

Ljung-Box Test

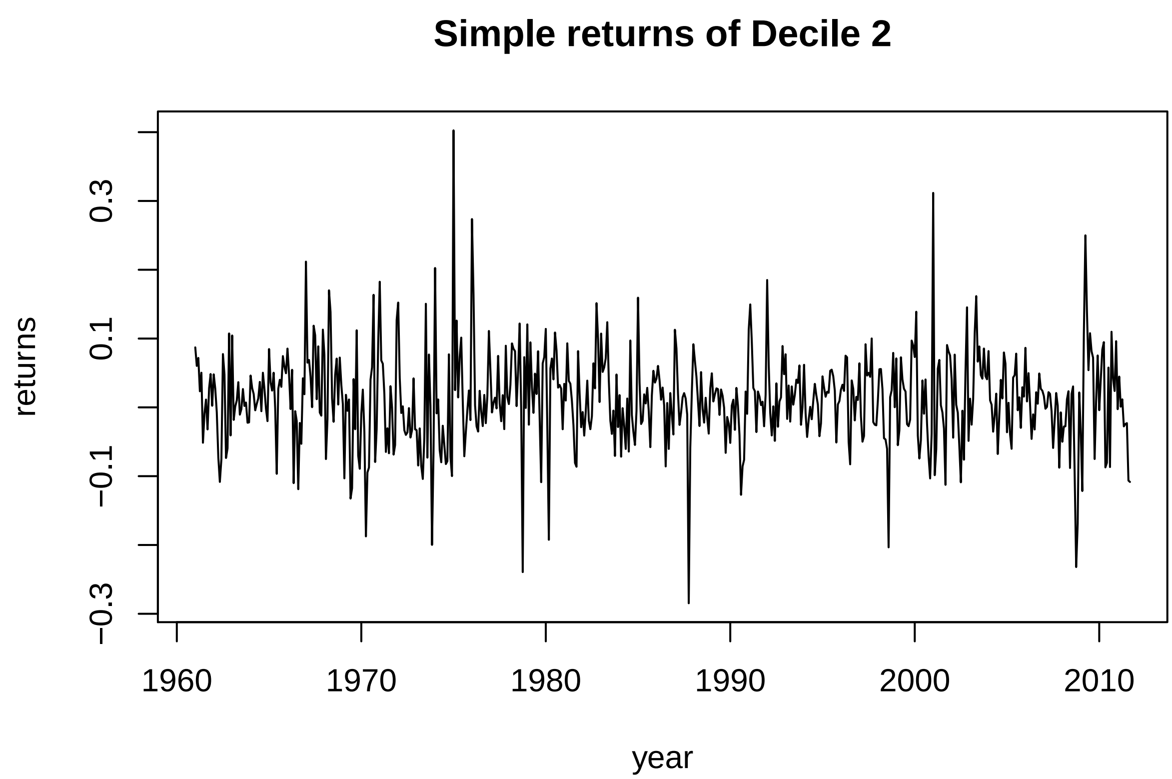
```
Box.test(d2.ts, lag=12, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  d2.ts
## X-squared = 17.176, df = 12, p-value = 0.1431
```

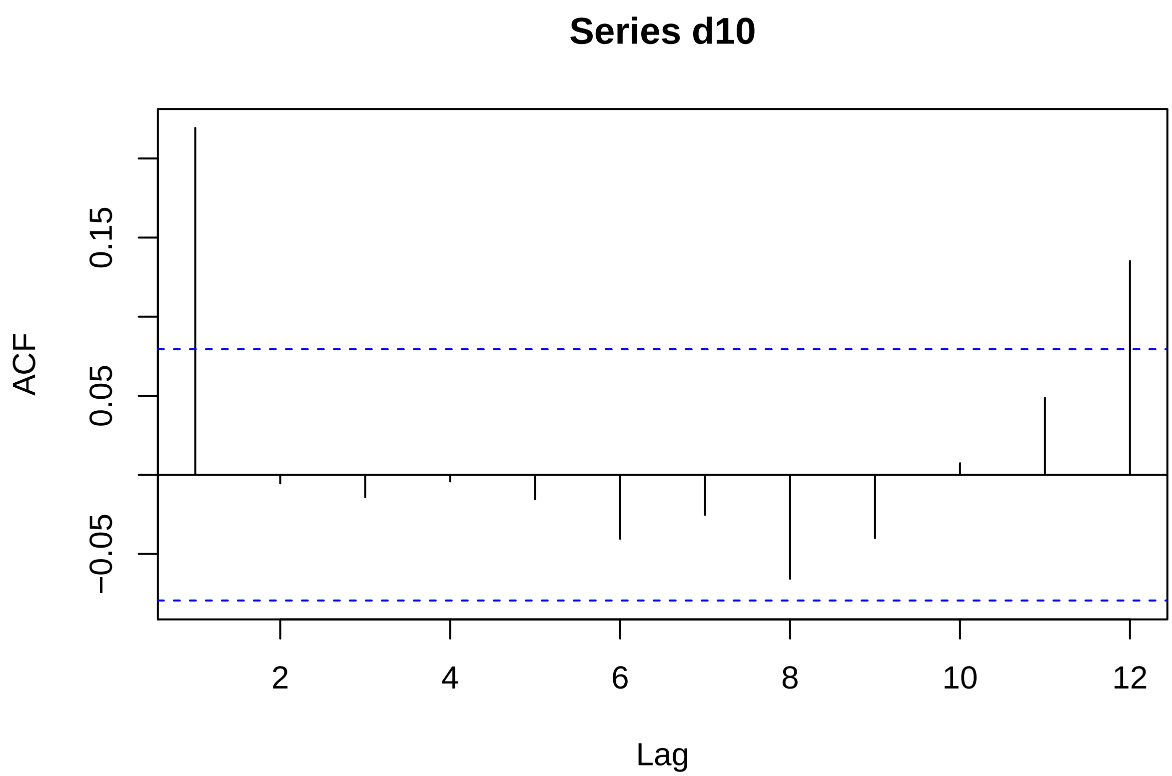
Based on Ljung-Box Test, p-values is larger that 0.05, We cannot reject the null hypothesis that all of the first 12 lags of autocorrelations are zero at the 5% level. There is an inconsistancy between Z-score and Ljung-Box Test. However, according to both ACF and PACF plot, we can clearly see an acorrelation of the first lag. We conclude that the first lag autocorrelation is significant enough to be non-zero.

Decile 10

```
# decile 10 simple returns
d10 <- dat$dec10
d10.ts <- ts(d10,frequency = 12, start=c(1961, 1))
# Plot the time series and its ACF for simple returns
plot(d10.ts, xlab="year", ylab="returns",type="l")
title(main = "Simple returns of Decile 2")
```



```r
f1=acf(d10,lag=12)
```




```r
tt=f1$acf*sqrt(length(d10))
critial.value=rep(qnorm(.975,0,1),length(tt))
compare=cbind(tt,qnorm(.975,0,1),tt>critial.value)
colnames(compare)=c('tt','critical value','tt > critial.value')
compare
```

Z-Score Test

```
##               tt critical value tt > critial.value
##  [1,]  5.4128049       1.959964                  1
##  [2,] -0.1330283       1.959964                  0
##  [3,] -0.3501727       1.959964                  0
##  [4,] -0.1029147       1.959964                  0
##  [5,] -0.3814595       1.959964                  0
##  [6,] -0.9974477       1.959964                  0
##  [7,] -0.6252607       1.959964                  0
##  [8,] -1.6217502       1.959964                  0
##  [9,] -0.9875511       1.959964                  0
## [10,]  0.1820250       1.959964                  0
## [11,]  1.2007877       1.959964                  0
## [12,]  3.3362712       1.959964                  1
```

Based on Z-scores test, both the first and 12th lag can reject the null hypothesis that autocorrelation equals zero with 5% level.

```r
Box.test(d10.ts, lag=12, type="Ljung")
```

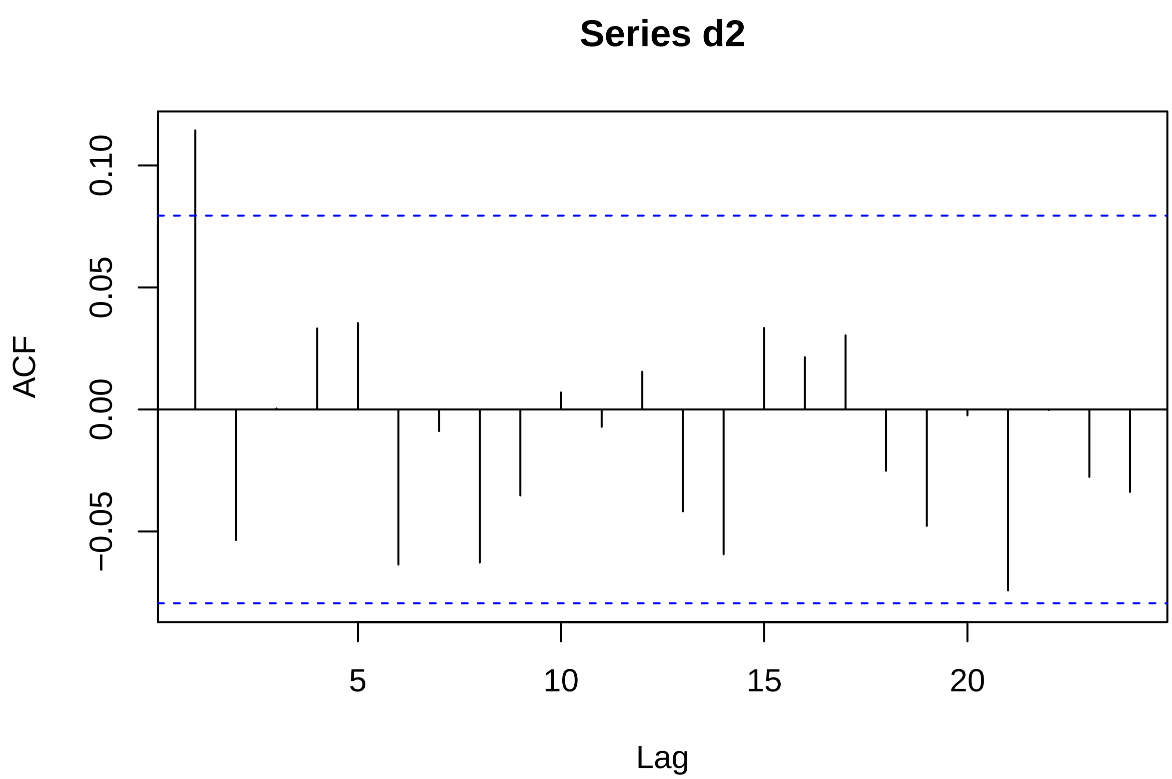
Ljung-Box Test

```
##
##  Box-Ljung test
##
## data:  d10.ts
## X-squared = 47.713, df = 12, p-value = 3.506e-06
```

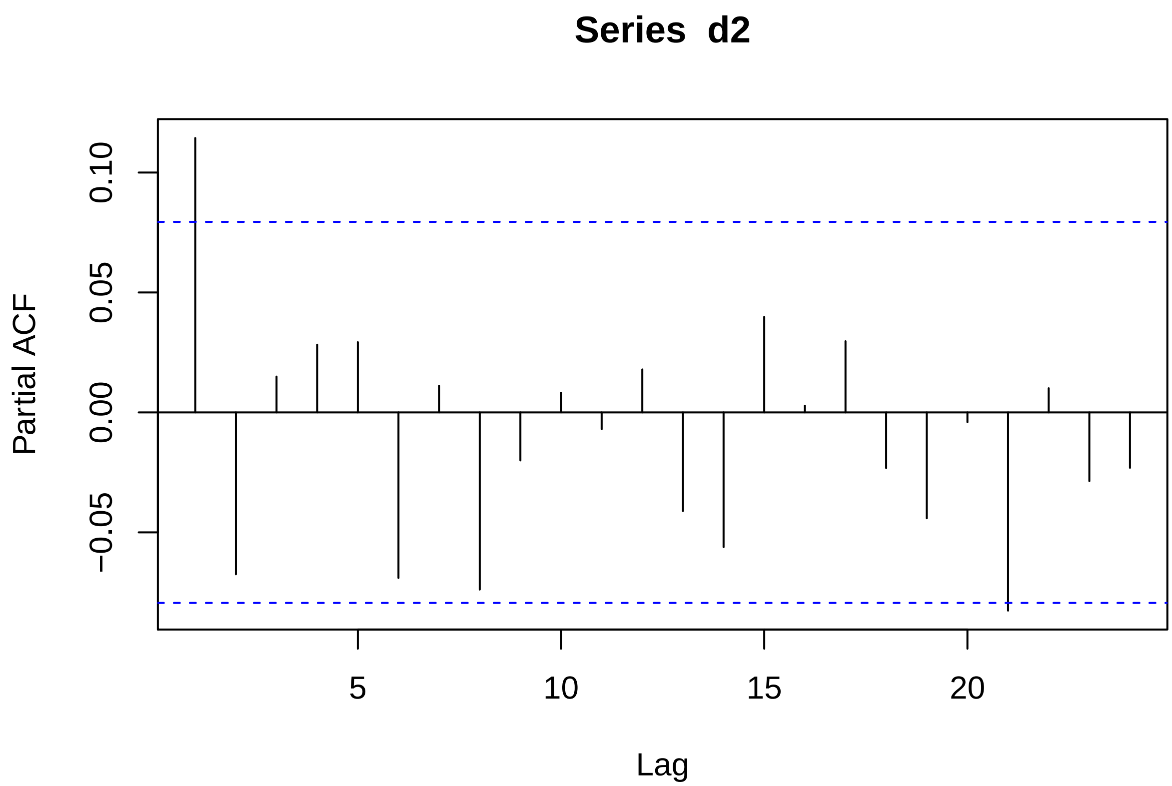
Based on Ljung-Box Test, p-values is smaller that 0.05, We can reject the null hypothesis that all of the first 12 lags of autocorrelations are zero at the 5% level. This conclusion is consistant with the Z-score test, since Z-score test rejected that lag 1 and 12 autocorrelations are zero at 5% level.

Fit ARMA model for returns of Decile2, perform model checking and write down the fitted model

```r
acf(d2, lag =24)
```



```r
pacf(d2, lag =24)
```



Based on ACF plot, we can see that MA(1) is a possible condidate for the model. PACF indicated that ARMA(1,1) could also be a model condidate. In addtion, PACF plot also indicated that there could be an autocorrelation for the 21st lag. It is too far into the future and only slightly above critical value. We could choose MA(1) or ARMA(1,1).

```r
eacf(d2.ts) # suggests the best fit is an MA(1)
```

```
## 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 o o o o o o o o o o  o  o  o
## 2 x x o o o o o o o o o  o  o  o
## 3 x o x o o o o o o o o  o  o  o
## 4 x o x o o o o o o o o  o  o  o
## 5 x x x x x o o o o o o  o  o  o
## 6 x x o x o x o o o o o  o  o  o
## 7 x x o x x x x o o o o  o  o  o
```

```r
auto.arima(d2.ts) # suggests the best fit is an MA(1)
```

```
## Series: d2.ts
## ARIMA(0,0,1) with non-zero mean
```

```
##
## Coefficients:
##          ma1    mean
##       0.1307  0.0093
## s.e.  0.0425  0.0022
##
## sigma^2 estimated as 0.00223:  log likelihood=996.04
## AIC=-1986.08   AICc=-1986.04   BIC=-1972.84
```

Both EACF and auto arima suggested MA(1) to be the best fit model.

```r
(ARMA11 <- arima(d2.ts, order=c(1,0,1)))
```

```
##
## Call:
## arima(x = d2.ts, order = c(1, 0, 1))
##
## Coefficients:
##           ar1     ma1  intercept
##       -0.4039  0.5265     0.0093
## s.e.   0.3885  0.3639     0.0021
##
## sigma^2 estimated as 0.002217:  log likelihood = 996.84,  aic = -1987.68
```

```r
(MA1 <- arima(d2.ts, order=c(0,0,1)))
```

```
##
## Call:
## arima(x = d2.ts, order = c(0, 0, 1))
##
## Coefficients:
##          ma1  intercept
##       0.1307     0.0093
## s.e.  0.0425     0.0022
##
## sigma^2 estimated as 0.002223:  log likelihood = 996.04,  aic = -1988.08
```

We can see that AIC for MA(1), -1988.08, is slightly lower than AIC of ARMA(1,1), -1987.68. MA(1) is a better choice.

Write down the model

ξ is the varible of the time series and α is the error.

B(X_t)=X_{t-1}

Equations for both models

ARMA(1,1): ξ_t=0.0093-0.4039B(ξ_t)+0.5265B(α_t)+α_t, σ_α^2=0.002217

MA(1): ξ_t=0.0093+0.1307B(α_t)+α_t, σ_α^2=0.002223

Note: MA(1) is our choice.

Model checking using Ljung-Box Test with Adjusted Degrees of Freedom.

Check if residuals behave like white noise using Box-Ljung test. (12 and 24 lags)

```
Box.test(MA1$residuals, lag=12, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  MA1$residuals
## X-squared = 9.4993, df = 12, p-value = 0.6598
```

```
Box.test(MA1$residuals, lag=24, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  MA1$residuals
## X-squared = 20.803, df = 24, p-value = 0.6503
```

Calculate adjusted p-value for the same statistic.

Adjested DF for 12 lags should be 12-1=11

Adjested DF for 24 lags should be 24-1=23

```
pv12=1-pchisq(9.4993,11) # Compute p value using 11 degrees of freedom
pv24=1-pchisq(20.803,23) # Compute p value using 23 degrees of freedom
paste(c(pv12,pv24))
```

```
## [1] "0.575905739010037" "0.593116770274183"
```

Bsed on the p-value calculated by DF adjusted Box-Ljung test, We cannot reject the hypothesis that all autocorrelations of lag residuals are equal to zero at 5% level. The residuals behave like white noise.

Use the fitted model to produce 1- to 12-step ahead forecasts of the series and the associated standard errors of forecasts.

```
MA1_train <- arima(d2.ts[1:597], order=c(0,0,1))
prd <- predict(MA1_train,12)
cbind(Actual=as.vector(tail(d2.ts,12)),
Predicted=as.vector(prd$pred),Standard.Error=as.vector(prd$se))
```

```
##          Actual   Predicted Standard.Error
##  [1,]  0.036242 0.021960199     0.04730190
##  [2,]  0.012035 0.009461795     0.04765859
##  [3,]  0.065768 0.009461795     0.04765859
##  [4,]  0.018406 0.009461795     0.04765859
##  [5,]  0.038320 0.009461795     0.04765859
##  [6,]  0.005935 0.009461795     0.04765859
```

```
##  [7,]  0.031454 0.009461795     0.04765859
##  [8,] -0.001103 0.009461795     0.04765859
##  [9,] -0.014384 0.009461795     0.04765859
## [10,] -0.036875 0.009461795     0.04765859
## [11,] -0.059275 0.009461795     0.04765859
## [12,] -0.080109 0.009461795     0.04765859
```

```
cbind(train.mean=mean(d2.ts[1:597]),train.sd=sd(d2.ts[1:597]))
```

```
##       train.mean   train.sd
## [1,] 0.009443459 0.04765428
```

It is clear that the predicted value converges to the mean of the time series and standard error converges to the standard deviation of the time series. The model eventually gives the unconditional mean value as a prediction. The quality of the forecast is very poor.

2. Exercise 4 on page 126

Consider the monthly yields of Moody's Aaa & Baa seasoned bonds from January 1919 to November, 2011. The data are obtained from FRED of Federal Reserve Bank of St. Louis. Consider the log series of monthly Aaa bond yields. Build a time series model for the series, including model checking.

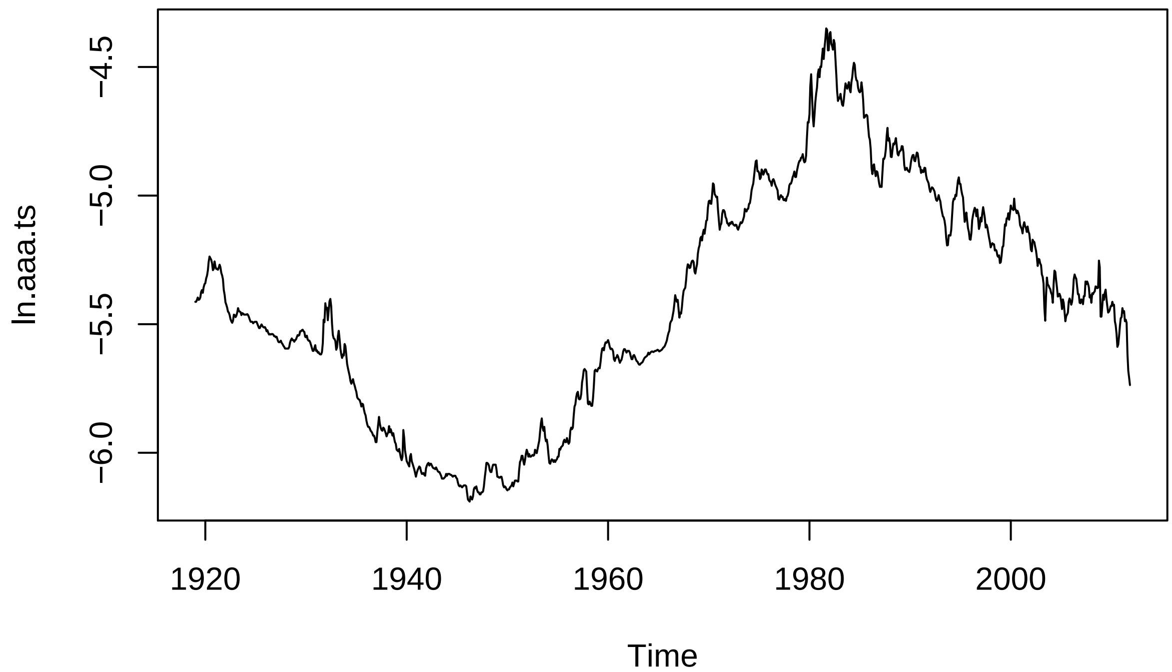
```
da=read.table(paste(datapath,"m-aaa-1911.txt",sep="/"),header=T)
tail(da)
```

```
##      year mon day yield
## 1110 2011   6   1  4.99
## 1111 2011   7   1  4.93
## 1112 2011   8   1  4.37
## 1113 2011   9   1  4.09
## 1114 2011  10   1  3.98
## 1115 2011  11   1  3.87
```

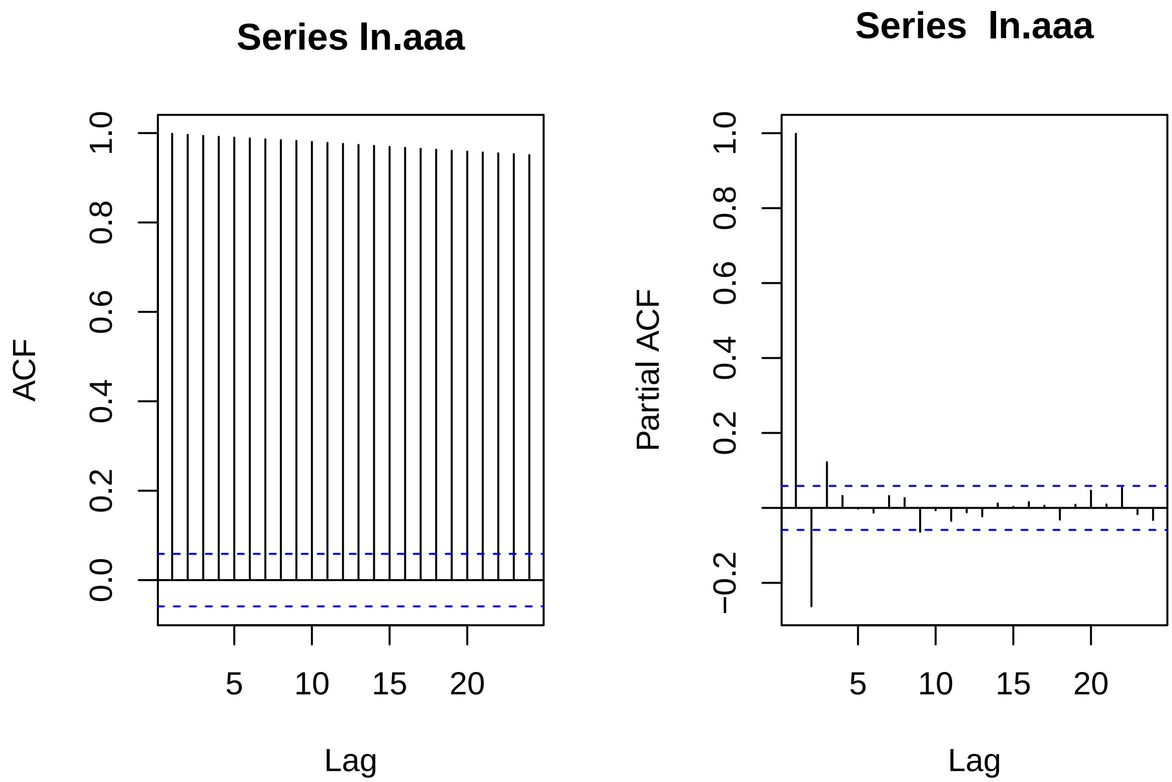
```
dim(da) #1115 4
```

```
## [1] 1115    4
```

```
ln.aaa<- log(da$yield/12/100)
# Build a time series
ln.aaa.ts <- ts(ln.aaa, frequency=12,start=c(1919,1))
plot(ln.aaa.ts)
```

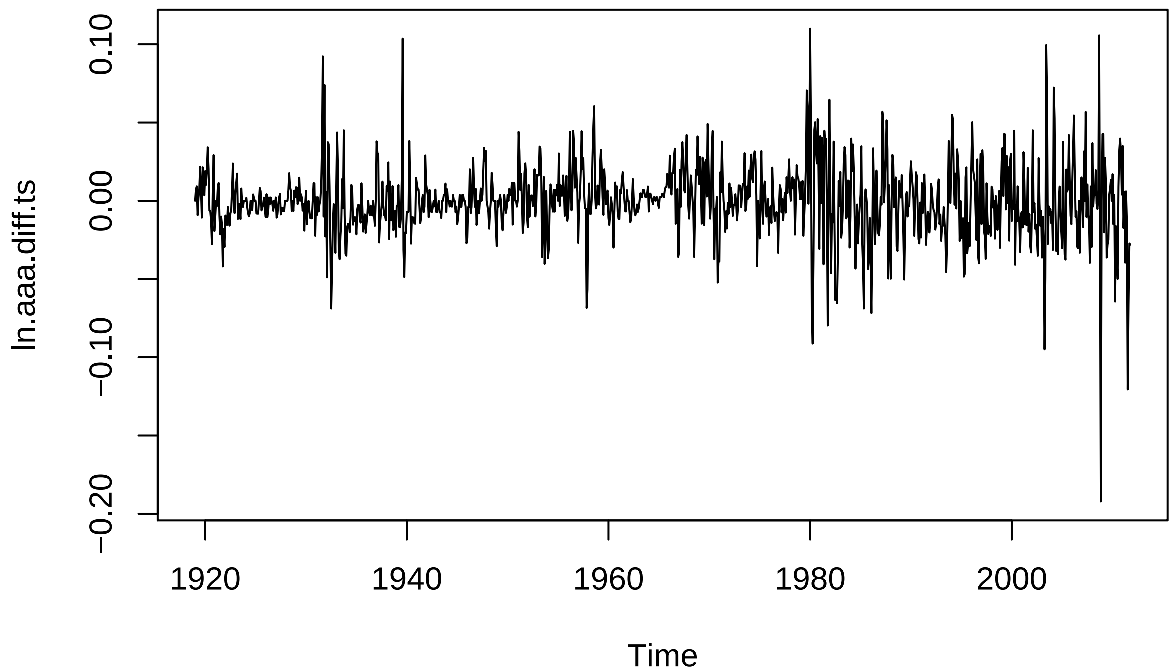



```
par(mfrow=c(1,2))
acf(ln.aaa, lag =24)
pacf(ln.aaa, lag =24)
```

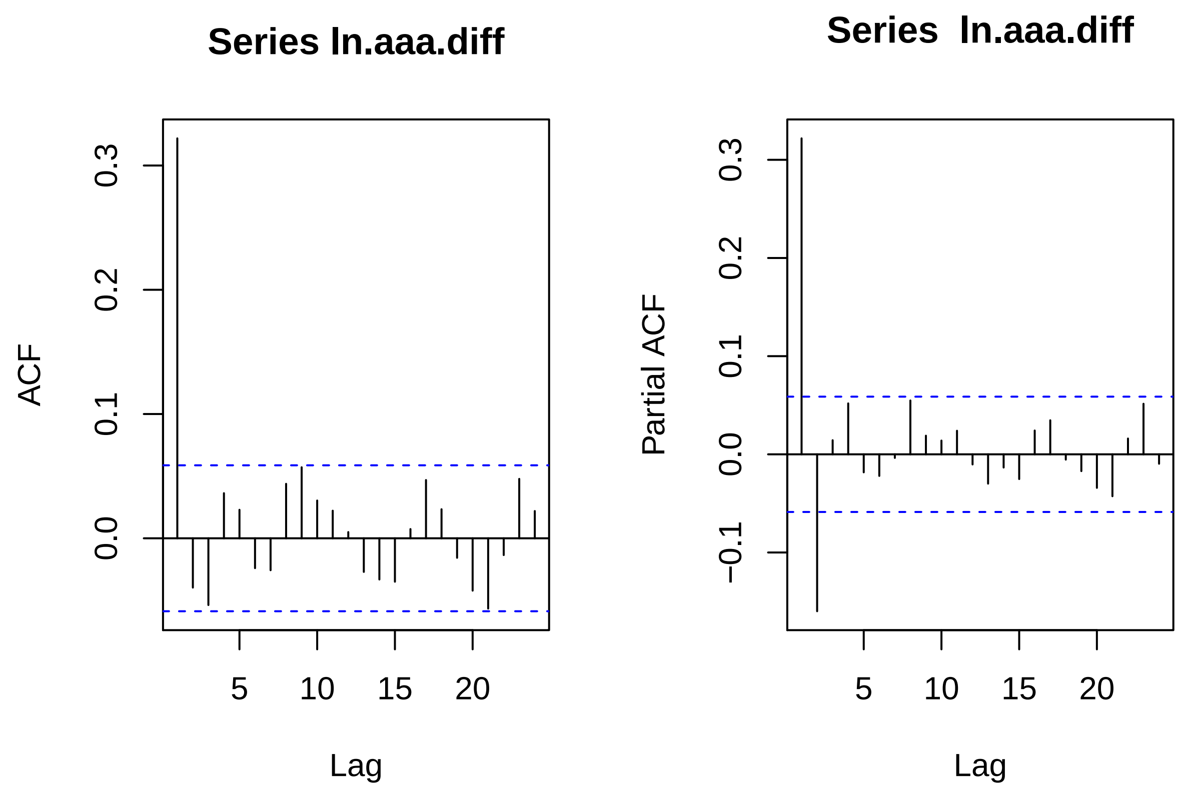


From both the time seires plot and ACF plots, we can see that the time series is not stationary. We need to take the first order differencing.

```
ln.aaa.diff<- diff(ln.aaa)
# Build a time series
ln.aaa.diff.ts <- ts(ln.aaa.diff, frequency=12,start=c(1919,1))
plot(ln.aaa.diff.ts)
```




```
par(mfrow=c(1,2))
acf(ln.aaa.diff, lag =24)
pacf(ln.aaa.diff, lag =24)
```



After the first order differencing, both ACF and Time series plots look stationary. Based on the ACF plot, we can see that first lag auto correlation is non-zero, so MA(1) can be a candidate. Based on PACF plot, we can see the first two PACs are non-zero, so ARMA(2,1) can also be a candidate.

```
# EACF suggests the best fit is MA(1)
eacf.ln.aaa.diff<-eacf(ln.aaa.diff.ts,6,12)$eacf
```

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

```
colnames(eacf.ln.aaa.diff)<-0:12
rownames(eacf.ln.aaa.diff)<-0:6
Compare.with<-2/sqrt(length(ln.aaa.diff.ts))
print(abs(eacf.ln.aaa.diff)-Compare.with,digits=2)
```

```
##       0     1       2      3       4      5      6      7       8      9     10
## 0 0.262 -0.02 -0.0062 -0.024 -0.0370 -0.036 -0.034 -0.016 -0.0028 -0.030 -0.038
## 1 0.346  0.19 -0.0149  0.013 -0.0380 -0.025 -0.046 -0.019 -0.0228 -0.047 -0.043
```

```
## 2 0.033  0.14  0.0171 -0.060 -0.0331 -0.051 -0.056 -0.018 -0.0130 -0.044 -0.052
## 3 0.198  0.24 -0.0099 -0.050 -0.0520 -0.037 -0.060 -0.033 -0.0088 -0.046 -0.045
## 4 0.258  0.32  0.0834 -0.041 -0.0524 -0.057 -0.042 -0.035 -0.0490 -0.054 -0.044
## 5 0.436  0.17  0.0675 -0.014 -0.0075  0.031 -0.059 -0.030 -0.0589 -0.054 -0.059
## 6 0.068  0.12  0.0785  0.294 -0.0511  0.014 -0.018 -0.030 -0.0593 -0.052 -0.057
##       11     12
## 0 -0.055 -0.033
## 1 -0.054 -0.041
## 2 -0.047 -0.041
## 3 -0.038 -0.058
## 4 -0.036 -0.056
## 5 -0.036 -0.052
## 6 -0.029 -0.050
```

```
# auto.arima suggests the best fit is also MA(1) (ARMA(0,1,1)) taking the first order differecing.
auto.arima(ln.aaa.ts)
```

```
## Series: ln.aaa.ts
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##       0.3695
## s.e.  0.0267
##
## sigma^2 estimated as 0.000467:  log likelihood=2691.48
## AIC=-5378.95   AICc=-5378.94   BIC=-5368.92
```

Both EACF and auto.arima suggest that MA(1) is the best model. Let take a look at both MA(1) and ARMA(2,1).

```
MA1.aaa<- arima(ln.aaa.ts, order=c(0,1,1))
ARMA21.aaa<- arima(ln.aaa.ts, order=c(2,1,1))
MA1.aaa
```

```
##
## Call:
## arima(x = ln.aaa.ts, order = c(0, 1, 1))
##
## Coefficients:
##          ma1
##       0.3695
## s.e.  0.0267
##
## sigma^2 estimated as 0.0004666:  log likelihood = 2691.48,  aic = -5380.95
```

```
ARMA21.aaa
```

```
##
## Call:
## arima(x = ln.aaa.ts, order = c(2, 1, 1))
##
```

```
## Coefficients:
##          ar1      ar2     ma1
##       0.3220  -0.1433  0.0531
## s.e.  0.1432   0.0549  0.1435
##
## sigma^2 estimated as 0.0004644:  log likelihood = 2694.08,  aic = -5382.15
```

```
cbind(MA1.AIC=AIC(MA1.aaa),
      MA1.BIC=AIC(MA1.aaa,k = log(length(sunspots))),
      ARMA21.AIC=AIC(ARMA21.aaa),
      ARMA21.BIC=AIC(ARMA21.aaa,k = log(length(sunspots))))
```

```
##        MA1.AIC   MA1.BIC ARMA21.AIC ARMA21.BIC
## [1,] -5378.951 -5367.062  -5380.153  -5356.375
```

We can see that ARMA(2,1) has a smaller AIC and MA(1) has a smaller BIC. It is hard to say which one is a better model fit.

Model checking using Ljung-Box Test with Adjusted Degrees of Freedom.

Check if residuals behave like white noise using Box-Ljung test. (12 and 24 lags)

Checking MA(1) Model

```
Box.test(MA1.aaa$residuals, lag=12, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  MA1.aaa$residuals
## X-squared = 14.619, df = 12, p-value = 0.263
```

```
Box.test(MA1.aaa$residuals, lag=24, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  MA1.aaa$residuals
## X-squared = 25.276, df = 24, p-value = 0.3909
```

Calculate adjusted p-value for the same statistic.

Adjested DF for 12 lags should be 12-1=11

Adjested DF for 24 lags should be 24-1=23

```
pv12.MA1=1-pchisq(14.619,11) # Compute p value using 11 degrees of freedom
pv24.MA1=1-pchisq(25.276,23) # Compute p value using 23 degrees of freedom
paste(c(pv12.MA1,pv24.MA1))
```

```
## [1] "0.200612002461827" "0.336213530098783"
```

Checking ARMA(2,1) Model

```r
Box.test(ARMA21.aaa$residuals, lag=12, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  ARMA21.aaa$residuals
## X-squared = 8.85, df = 12, p-value = 0.7157
```

```r
Box.test(ARMA21.aaa$residuals, lag=24, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  ARMA21.aaa$residuals
## X-squared = 18.448, df = 24, p-value = 0.7807
```

Calculate adjusted p-value for the same statistic.

Adjested DF for 12 lags should be 12-3=9

Adjested DF for 24 lags should be 24-3=21

```r
pv12.ARMA21=1-pchisq(8.85,9) # Compute p value using 11 degrees of freedom
pv24.ARMA21=1-pchisq(18.448,21) # Compute p value using 23 degrees of freedom
paste(c(pv12.ARMA21,pv24.ARMA21))
```

```
## [1] "0.451234288991281" "0.620504014066719"
```

Based on the p-value calculated by DF adjusted Box-Ljung test, We cannot reject the hypothesis that all autocorrelations of lag residuals are equal to zero at 5% level for either model. The residuals of both models behave like white noises.