Exercise 5

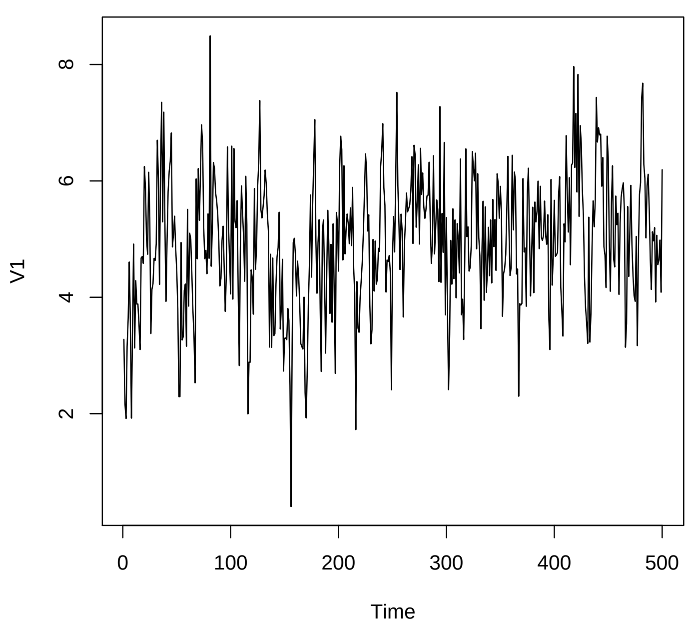
Aliaksandr Panko, Nadzeya Yakimchyk, Nikolai Agafonov, Ilia Kurennoi

June 21, 2018

```
> library("tseries")
> library("forecast")
> library("urca")
> seriesA <-read.csv("seriesA.csv",head=FALSE,sep = "")
> seriesA <- as.ts(seriesA)
> seriesB <-read.csv("seriesB.csv",head=FALSE,sep = "")
> seriesB <- as.ts(seriesB)
```

Let's start with seriesA. The graph shows that it has good chances to be stationary:

```
> plot(seriesA)
```



Now let's perform the augmented Dickey-Fuller (ADF) test to check whether

there is unitroot.
Enders, Applied Econometric Time Series, 2nd edition, p. 213 advices to start with the least restrictive of the plausible models (which will generally include a trend and drift) and use tau3 statistic to test the null hypothesis that $\gamma = 0$. Unit root tests have low power to reject the null hypothesis; hence, if the null hypothesis of a unit root is rejected, there is no need to proceed further. So let's do it and use AIC to select lags:

```
> summary(urca::ur.df(seriesA,type="trend",lags=7,selectlags = "AIC"))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3842 -0.7039 -0.0076  0.6761  3.6649 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.2201079  0.2350361   9.446  < 2e-16 ***
z.lag.1     -0.4782790  0.0472188 -10.129  < 2e-16 ***
tt           0.0006163  0.0003210   1.920 0.055436 .  
z.diff.lag  -0.1597182  0.0445621  -3.584 0.000372 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9886 on 488 degrees of freedom
Multiple R-squared:  0.305,	Adjusted R-squared:  0.3007 
F-statistic: 71.38 on 3 and 488 DF,  p-value: < 2.2e-16


Value of test-statistic is: -10.129 34.2276 51.3206 

Critical values for test statistics: 
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

Test statistic -10.129 is lower than critical values for any significance level con-

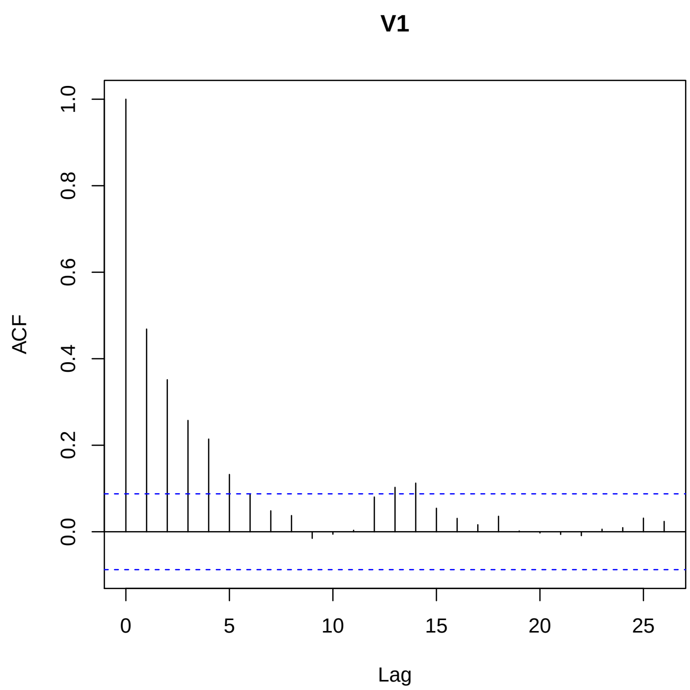
sidered, so we can reject the null and accept the alternative that the data is stationary.
Basically, the same conclusion can be drawn if we use ADF test from *tseries* package with default settings (as we have seen on the lecture), which with more details can be replicated with *urca* package if not use AIC to select lags (so test statistic is -7.0616, but still much lower than critical values):

```
> tseries::adf.test(seriesA)

        Augmented Dickey-Fuller Test

data:  seriesA
Dickey-Fuller = -7.0616, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

> summary(urca::ur.df(seriesA,type="trend",lags=7))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4325 -0.6923 -0.0116  0.6566  3.7466

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.2476900  0.3252947   6.910 1.54e-11 ***
z.lag.1     -0.4850200  0.0686837  -7.062 5.78e-12 ***
tt           0.0006365  0.0003326   1.914   0.0562 .
z.diff.lag1 -0.1621168  0.0696224  -2.329   0.0203 *
z.diff.lag2 -0.0219793  0.0671943  -0.327   0.7437
z.diff.lag3  0.0201491  0.0642685   0.314   0.7540
z.diff.lag4  0.0816736  0.0615566   1.327   0.1852
z.diff.lag5  0.0498713  0.0578460   0.862   0.3890
z.diff.lag6  0.0253625  0.0533474   0.475   0.6347
z.diff.lag7 -0.0021464  0.0451561  -0.048   0.9621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9908 on 482 degrees of freedom
```

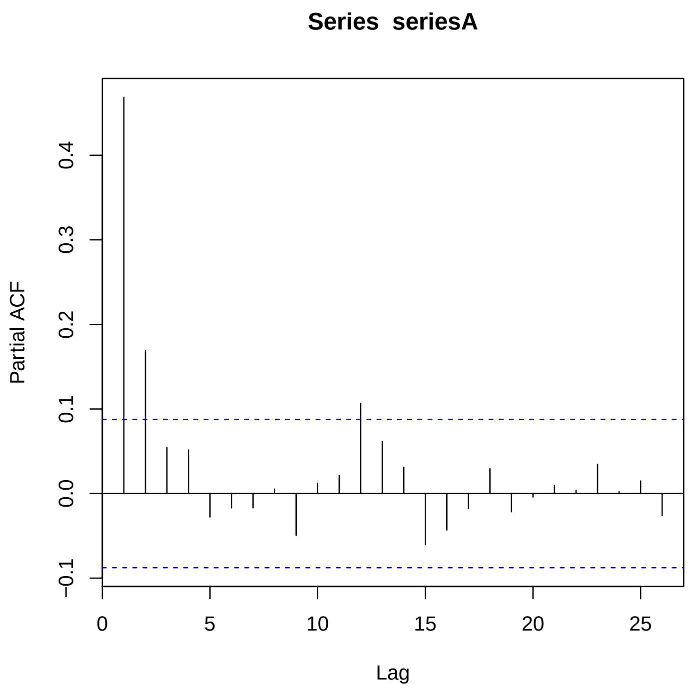
```
Multiple R-squared:  0.3105,	Adjusted R-squared:  0.2976
F-statistic: 24.12 on 9 and 482 DF,  p-value: < 2.2e-16


Value of test-statistic is: -7.0616 16.6976 25.0088

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

So we can use seriesA "as is" and fit ARMA model. Let's examine autocorrelations and partial autocorrelations:

```
> acf(seriesA)
```




```
> pacf(seriesA)
```

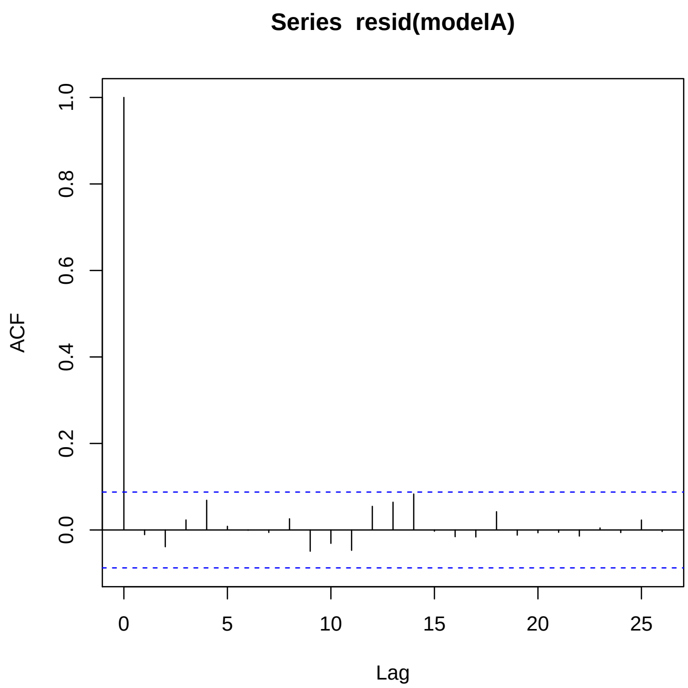


Partial autocorrelations cut off at lag 2, so we take p=2. Autocorrelations show exponential decay (oscillating). So let's fit the model AR(2):

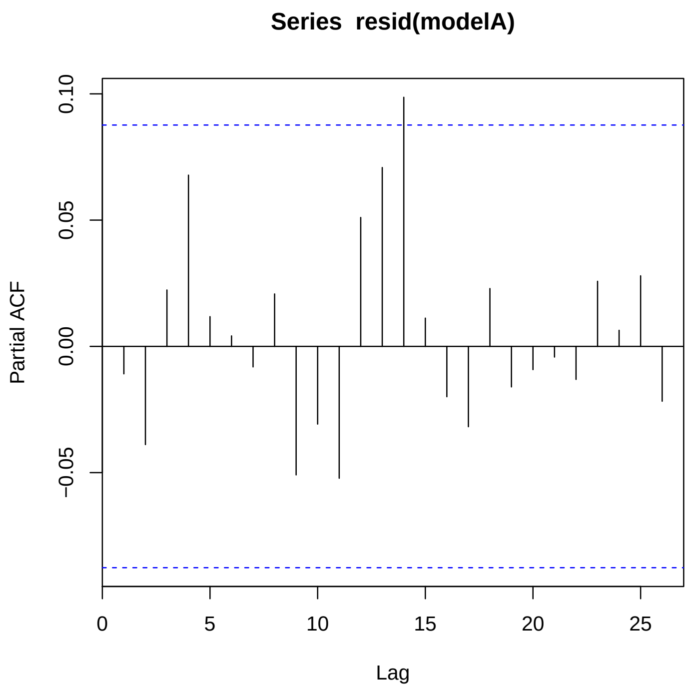
```
> modelA <- arima(seriesA, order = c(2,0,0))
> summary(modelA)

Call:
arima(x = seriesA, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.3899  0.1731     4.9241
s.e.  0.0441  0.0444     0.1016

sigma^2 estimated as 0.9934:  log likelihood = -707.97,  aic = 1423.95

Training set error measures:
                      ME      RMSE       MAE       MPE     MAPE      MASE
Training set 0.003569965 0.9967012 0.7985911 -6.376913 19.73943 0.8476101
                    ACF1
Training set -0.01088213
```

The graphs show that there is no autocorrelation:

```
> acf(resid(modelA))
```




```
> pacf(resid(modelA))
```

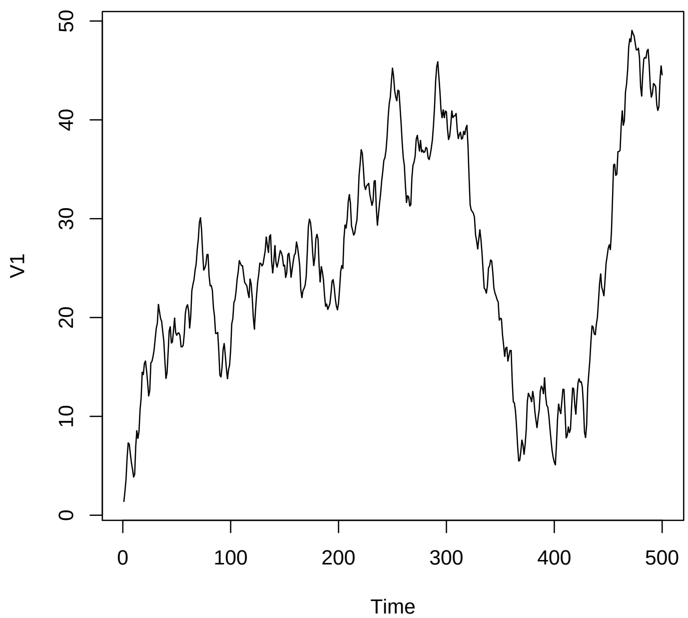


And it is confirmed by Ljung-Box test:

```
> Box.test(resid(modelA),type="Ljung",lag=20,fitdf=1)

        Box-Ljung test

data:  resid(modelA)
X-squared = 15.246, df = 19, p-value = 0.7068
```

Now let's analyize series B.
The graph shows that it is likely to be nonstationary:

```
> plot(seriesB)
```



Let's run the test:

```
> summary(urca::ur.df(seriesB,type="trend",lags=7,selectlags = "AIC"))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2229 -0.7481 -0.0600  0.7081  3.2150

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2326170  0.1350629   1.722   0.0857 .
z.lag.1     -0.0083819  0.0045060  -1.860   0.0635 .
tt           0.0001184  0.0003450   0.343   0.7316
```

```
z.diff.lag1  0.6388427  0.0451310  14.155  < 2e-16 ***
z.diff.lag2 -0.3310402  0.0514403  -6.435 2.96e-10 ***
z.diff.lag3  0.0812189  0.0453870   1.789   0.0742 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.054 on 486 degrees of freedom
Multiple R-squared:  0.3005,	Adjusted R-squared:  0.2933
F-statistic: 41.75 on 5 and 486 DF,  p-value: < 2.2e-16


Value of test-statistic is: -1.8602 1.5066 1.7365

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

-1.8602 is larger than critical values for tau3, so the null hypothesis that there is a unit root is accepted.
1.7365 is less than critical values for phi3, so the null hypothesis that there is a unit root AND no-trend (without trend) is accepted.
1.5066 is less than critical values for phi2, so the null hypothesis hypothesis that there is a unit root without trend and without drift is accepted.
So we conclude that seriesB is not stationary.
To run ARMA let's make series B stationary by taking differences:

```
> difB <- diff(seriesB)
```

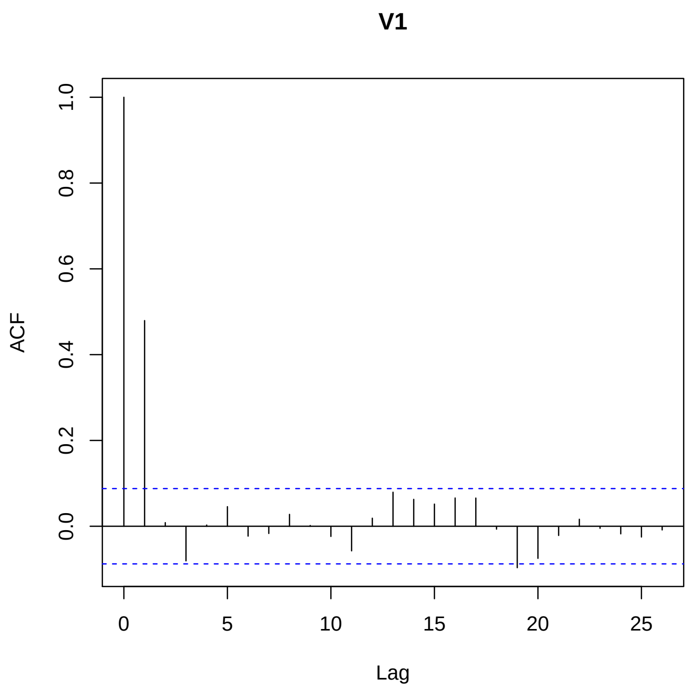
Let's run the test again:

```
> summary(urca::ur.df(difB,type="trend",lags=7,selectlags = "AIC"))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1036 -0.7290 -0.0534  0.7291  3.2221
```

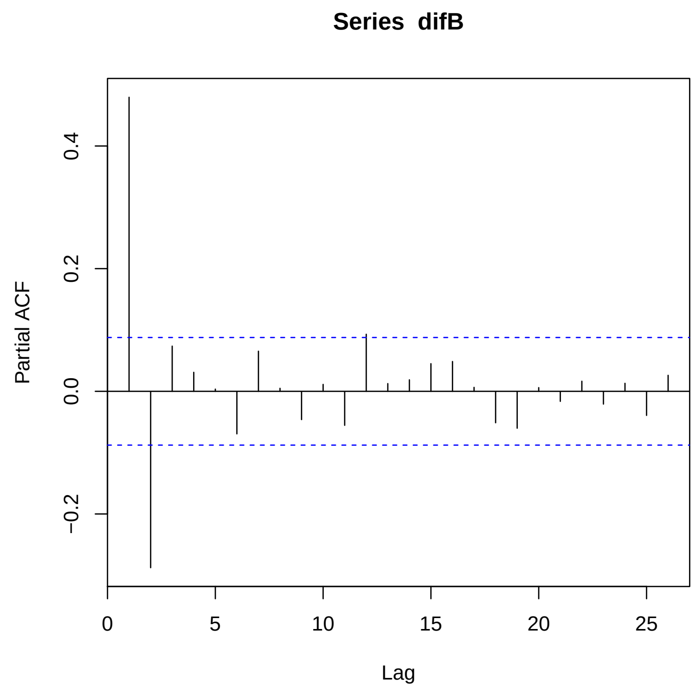
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.273e-02  9.776e-02   0.642    0.521
z.lag.1     -6.242e-01  5.379e-02 -11.603  < 2e-16 ***
tt          -5.075e-05  3.367e-04  -0.151    0.880
z.diff.lag1  2.606e-01  4.686e-02   5.561 4.43e-08 ***
z.diff.lag2 -7.555e-02  4.543e-02  -1.663    0.097 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.058 on 486 degrees of freedom
Multiple R-squared:  0.3273,        Adjusted R-squared:  0.3218
F-statistic: 59.12 on 4 and 486 DF,  p-value: < 2.2e-16


Value of test-statistic is: -11.6032 44.8814 67.3206

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

So difB is stationary.

```
> acf(difB)
```




```
> pacf(difB)
```



The graphs of autocorrelations and partial autocorrelations do not really fit patterns which are described in the table in lecture notes, p. 103. We use the following function to choose the model:

```
>   forecast::auto.arima(difB, seasonal=FALSE)

Series: difB
ARIMA(3,0,0) with zero mean

Coefficients:
         ar1      ar2     ar3
      0.6434  -0.3351  0.0758
s.e.  0.0447   0.0510  0.0449

sigma^2 estimated as 1.109:  log likelihood=-732.56
AIC=1473.12   AICc=1473.2   BIC=1489.97
```

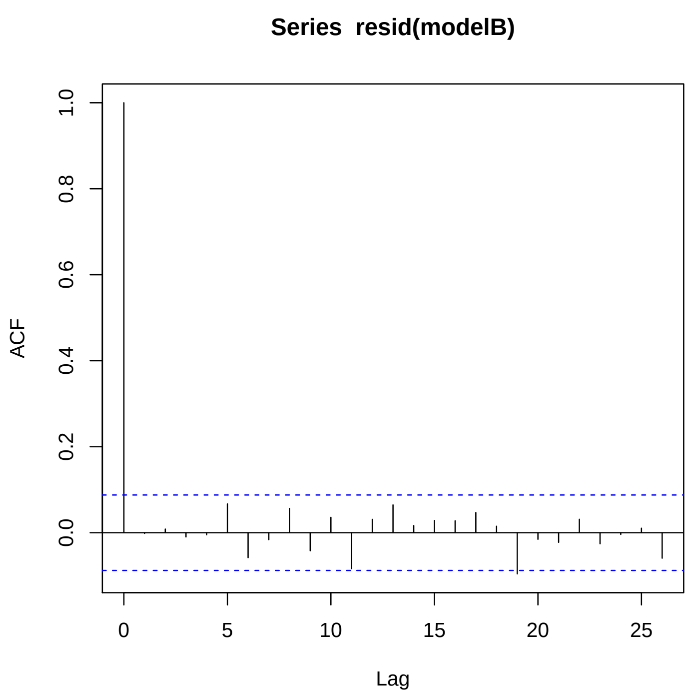
Let's fit the model:

```
> modelB <- arima(difB, order = c(3,0,0))
> summary(modelB)

Call:
arima(x = difB, order = c(3, 0, 0))
```

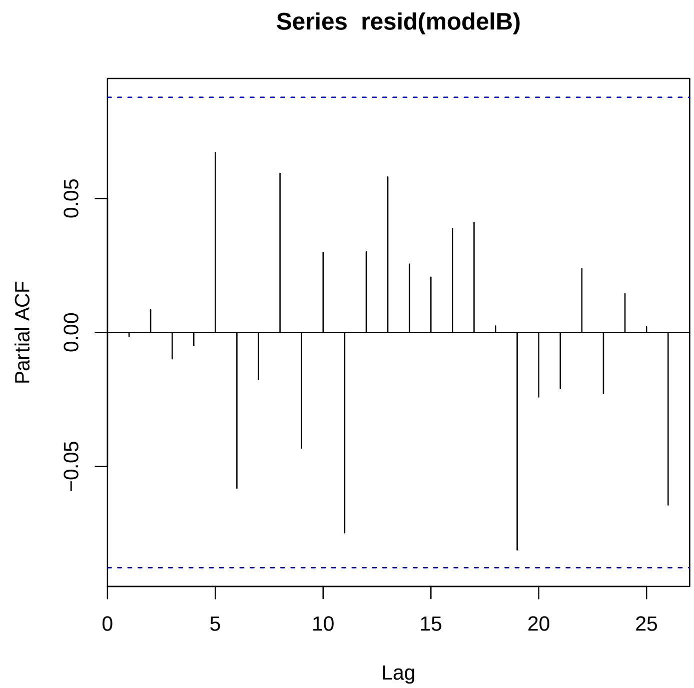
```
Coefficients:
         ar1      ar2     ar3  intercept
      0.6406  -0.3354  0.0736     0.0854
s.e.  0.0447   0.0510  0.0449     0.0755

sigma^2 estimated as 1.099:  log likelihood = -731.93,  aic = 1473.86

Training set error measures:
                        ME     RMSE       MAE      MPE     MAPE      MASE
Training set -0.0005370063 1.048537 0.8488529 581.5029 710.7113 0.8295792
                     ACF1
Training set -0.001529711
```

The graphs show that there is no autocorrelation:

> acf(resid(modelB))



> pacf(resid(modelB))



And it is confirmed by Ljung-Box test:

```
> Box.test(resid(modelB),type="Ljung",lag=20,fitdf=1)

        Box-Ljung test

data:  resid(modelB)
X-squared = 20.733, df = 19, p-value = 0.3517
```