Financial Econometrics Rauli Susmel
FINA 4397

Homework 3 - Solutions

1.1

a) MA(1)
AR(1)
RW with drift

b)
ACF for MA(1): Non-zero for first autocorrelation, zero for 2nd and beyond autocorrelations
ACF(1) = 0.5
ACF(k) = 0 for k>1

ACF for AR(1): Tail-off
ACF(1) = 0.8
ACF(2) = 0.8^2
ACF(3) = 0.8^3

ACF for RW: Constant and equal to 1.
ACF(k) = 1 for k= 1, 2, 3, ...

c) RW with drift. Unpredictable returns
Predicting the stock market best: The AR(1) model.

d) MA(1): Impact of shock lasts only one period.
AR(1): Impact of shock last a long time, but impact decays at rate 0.8.
RW: Impact of shock last forever.

1.2

a. To check for stationarity, we look at the roots of $\phi(z)$:
The roots of $\phi(z) = 1 - .803\, z - .682\, z^2$ should lie outside the unit circle.
Let's use R:

```
x1 <- c(1, -.803, -.682)          # define the coefficients of polynomial
> polyroot(x1)
[1]  0.7577133+0i  -1.9351326-0i
```

Conclusion: We have two roots: z_1 = **0.7577133**; z_2 = **-1.9351326**. One is inside (z_1), one is outside (z_2) $\Rightarrow$ non-stationary process

b. Constant moments. The distribution does not change with time.

c. Patterns
ACF: Tailing-off at a low rate.
PACF: Cut-off after lag 2

d. Produce one-, two-, and three-step ahead forecasts.

$$\hat{Y}_{t+1} = .803\, Y_t + .682\, Y_{t-1}$$

$\hat{Y}_{t+2} = .803\,\hat{Y}_{t+1} + .682\,Y_t$
$\hat{Y}_{t+3} = .803\,\hat{Y}_{t+2} + .682\,\hat{Y}_{t+1}$

1.3

a. SE = 1/sqrt(500) = **0.04472136** $\Rightarrow$ Significant ACF: 1, 5
Significant PACF: 1, 2, 3

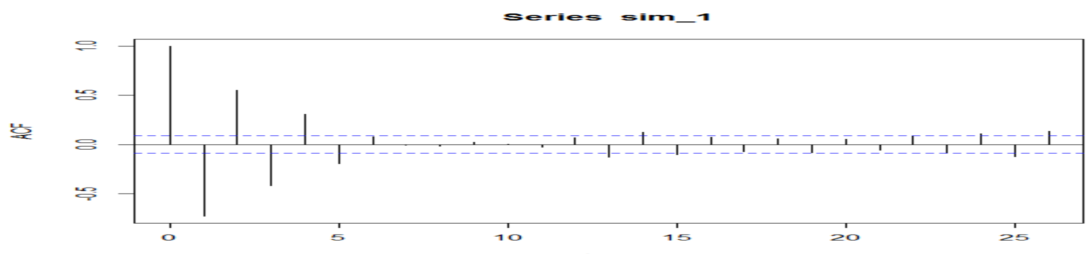
b. LB(5) = 500 * (500+2) * (0.307^2/(500-1) + 0.013^2/(500-2) + 0.086^2/(500-3) + 0.031 ^2/(500-4) + 0.197^2/(500-5)) = **71.39342** $\Rightarrow$ very significant at 5% level. That is, we reject H_0: all five autocorrelations are jointly zero.

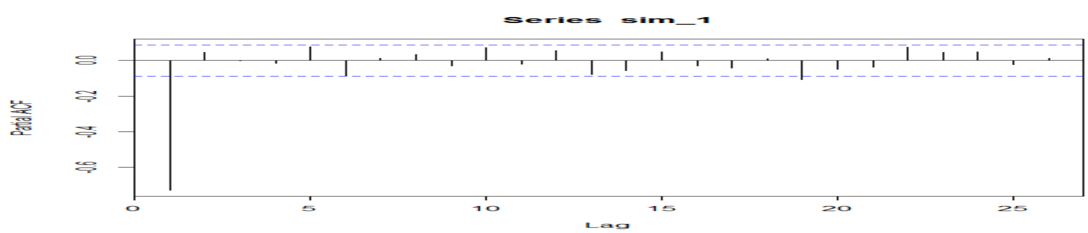
c. PACFs are tailing off, MA process likely. With the information above, possible MA(5) process.

1.4

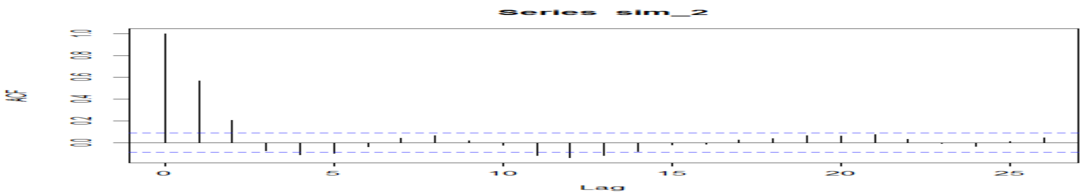
a. Plot the ACF/PACF for each series.
SIM_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/Sim.arima.csv", head=TRUE, sep=",")
sim_1 <-SIM_da$sim_1
sim_2 <-SIM_da$sim_2
sim_3 <-SIM_da$sim_3
sim_4 <-SIM_da$sim_4

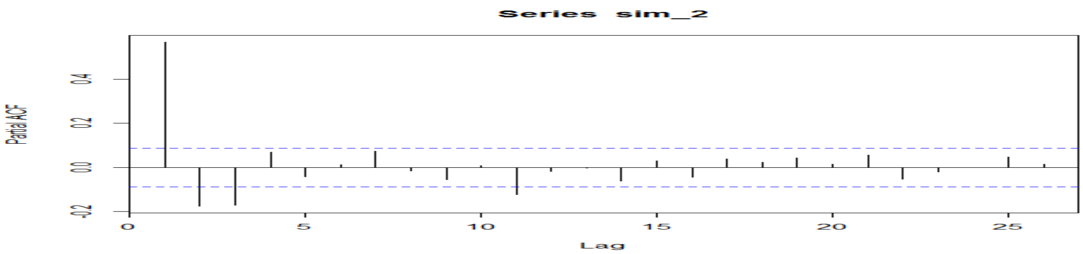
Sim_1
acf(sim_1)
pacf(sim_1)



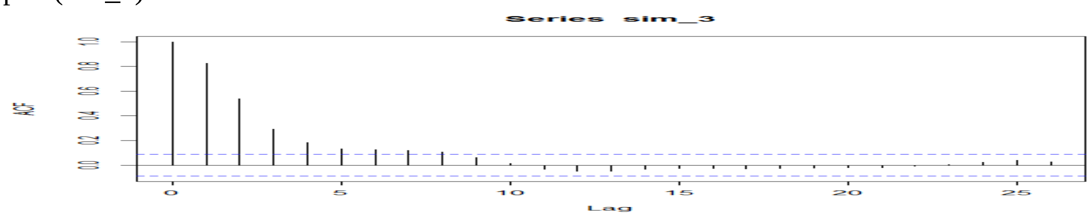


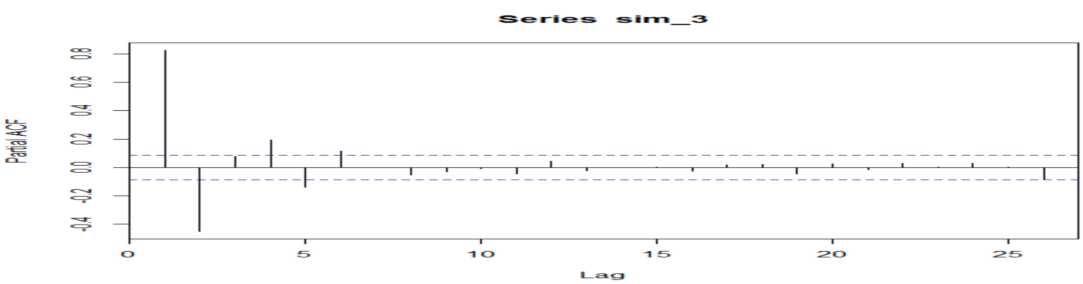
Sim_2
acf(sim_2)
pacf(sim_2)



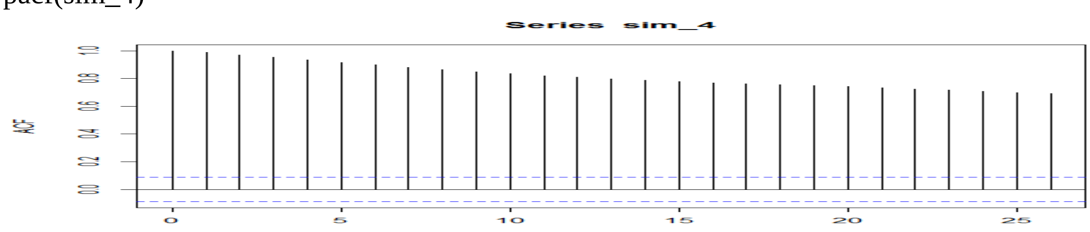


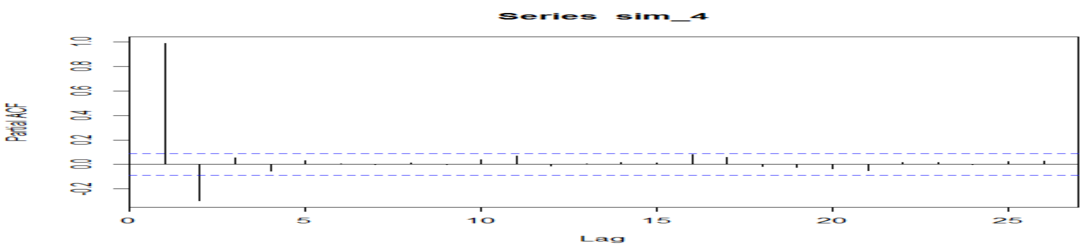
Sim_3
acf(sim_3)
pacf(sim_3)





Sim_4
acf(sim_4)
pacf(sim_4)





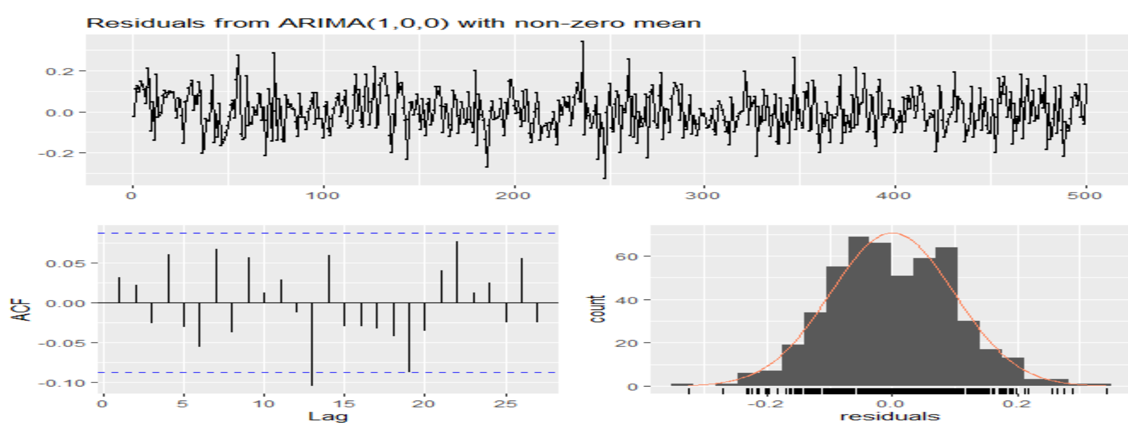
b. Using the ACF/PACF guess the order of the ARIMA process
Sim_1: AR(1)
Sim_2: MA(2)
Sim_3: ARMA(1,2)
Sim_4: ARIMA(0,1,1)

c. Check the residuals to see if there is any autocorrelation left in each of your series.
Sim_1

```
fit_1 <- arima(sim_1, order=c(1,0,0))
> fit_1
Coefficients:
         ar1  intercept
     -0.7307    -0.0007
s.e.  0.0305     0.0026

sigma^2 estimated as 0.009822:  log likelihood = 445.93,  aic = -885.85

> checkresiduals(fit_1)
```



Repeat for the other series.

1.5 (Forecasting with ARMA and SES)
a. Forecasts. We use the formula:

$$\hat{Y}_{T+l} = 0.036 + .69\, \hat{Y}_{T+l-1} + 0.42\, \mathrm{E}[\varepsilon_{T+\ell-1}|IT] + \mathrm{E}\left[\varepsilon_{T+\ell-q}|IT\right]$$

$$\hat{Y}_T = 0.036 + .69 * 3.4 + 0.42 * (-1.3) = 1.836$$

$\hat{Y}_{T+1} = 0.036 + .69 * 1.836 = 1.30284$
$\hat{Y}_{T+2} = 0.036 + .69 * 1.30284 = 0.9349596$

b. MSE_arma = ((1.836 - (-0.032))^2 + (1.30284 - 0.961)^2 + (0.9349596 - 0.203)^2)/3 =
= **1.380681**

c. $S_t = S_{t-1} + .15 * (Y_{t-1} - S_{t-1})$
$\hat{Y}_T = S_T = 0.0305 + .15 * (3.4 - 0.0305) = 0.535925$
$\hat{Y}_{T+1} = 0.535925$
$\hat{Y}_{T+2} = 0.535925$

MSE_ses = ((0.535925 - (-0.032))^2 + (0.535925- 0.961)^2 +(0.535925 - 0.203)^2)/3 =
= **0.2046889**

d. SES has lower MSE => Preferred model.

1.6 (Forecasting Evaluation)
a.
A: $\hat{x}_{T+1} = 0.38 + .10 * (-0.02) = 0.378$
$\hat{x}_{T+2} = 0.38$
$\hat{x}_{T+3} = 0.38$
$\hat{x}_{T+4} = 0.38$

$\hat{Y}_{T+l} = 0.38 + .10\, \hat{Y}_{T+l-1} - 0.09\, \hat{Y}_{T+l-2}$

B: $\hat{x}_{T+1} = 0.63 + .17 * 0.31 - .009 * 0.02 = 0.68252$
$\hat{x}_{T+2} = 0.63 + .17 * 0.68252 - .009 * 0.31 = 0.7432384$
$\hat{x}_{T+3} = 0.63 + .17 * 0.7432384 - .009 * 0.68252 = 0.7502078$
$\hat{x}_{T+4} = 0.63 + .17 * 0.7502078 - .009 * 0.7432384 = 0.7508462$

b.
MAE_ma1 = (|0.378 - (0.62)| + |0.38 – (0.19)| + |0.38 – (-0.32)|^2 + |0.38 – (0.72)|)/4
= **0.368**

MAE_ar2 = (|0.68252 - (0.62)| + |0.7432384 - 0.19| + |0.7502078 - (-0.32)| + + |0.7508462 – (0.72)|)/4 = **0.4292**

Conclusion: MA model has better performance.

1.7
a.
Sh_da <- read.csv("C:/Financial Econometrics/Real_Estate_2019.csv",head=TRUE,sep=",")

x_dat <- Sh_da$DATE
x_sf1 <- Sh_da$SF_c
x_sf <- x_lv[1:322]

```
fit_ar3 <- arima(x_sf, order=c(3,0,0))
> fit_ar3

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

Coefficients:
         ar1     ar2      ar3  intercept
      1.0141  0.0508  -0.1791     0.0029
s.e.  0.0546  0.0787   0.0546     0.0022

sigma^2 estimated as 2.175e-05:  log likelihood = 1270.64,  aic = -2531.29
```

b.

```
T_sf <- length(res_sf_ar3)
Feb <-rep(c(1,0,0,0,0,0,0,0,0,0,0,0),T_lv/12+1) # Create January dummy
Mar <-rep(c(0,1,0,0,0,0,0,0,0,0,0,0),T_lv/12+1) # Create February dummy
Apr <-rep(c(0,0,1,0,0,0,0,0,0,0,0,0),T_lv/12+1)
May <-rep(c(0,0,0,1,0,0,0,0,0,0,0,0),T_lv/12+1)
Jun <-rep(c(0,0,0,0,1,0,0,0,0,0,0,0),T_lv/12+1)
Jul <-rep(c(0,0,0,0,0,1,0,0,0,0,0,0),T_lv/12+1)
Aug <-rep(c(0,0,0,0,0,0,1,0,0,0,0,0),T_lv/12+1)
Sep <-rep(c(0,0,0,0,0,0,0,1,0,0,0,0),T_lv/12+1)
Oct <-rep(c(0,0,0,0,0,0,0,0,1,0,0,0),T_lv/12+1)
Nov <-rep(c(0,0,0,0,0,0,0,0,0,1,0,0),T_lv/12+1)
Dec <-rep(c(0,0,0,0,0,0,0,0,0,0,1,0),T_lv/12+1)
seas_sf <-cbind(Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec)
seas <- seas_sf[1:(T_sf),]
fit_sf_seas <- lm(res_sf_ar3 ~ seas)
> summary(fit_sf_seas)

Call:
lm(formula = res_sf_ar3 ~ seas)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0154680 -0.0025181  0.0000796  0.0026429  0.0118629

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0009787  0.0008446   1.159  0.24744
seasFeb     -0.0011926  0.0011834  -1.008  0.31433
seasMar      0.0025801  0.0011834   2.180  0.02999 *
seasApr      0.0030514  0.0011834   2.579  0.01038 *
seasMay     -0.0002333  0.0011834  -0.197  0.84387
seasJun     -0.0009945  0.0011834  -0.840  0.40136
seasJul     -0.0029739  0.0011834  -2.513  0.01248 *
seasAug     -0.0020289  0.0011834  -1.714  0.08745 .
seasSep     -0.0035327  0.0011834  -2.985  0.00306 **
seasOct     -0.0024083  0.0011834  -2.035  0.04269 *
seasNov     -0.0024161  0.0011834  -2.042  0.04203 *
```

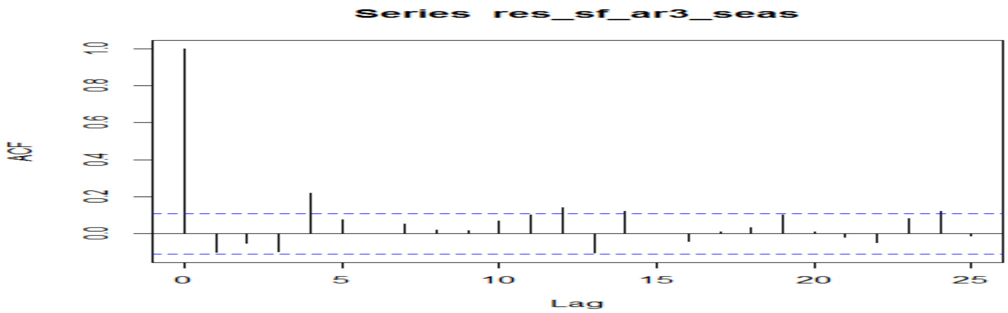
seasDec   -0.0014500 0.0011945 -1.214 0.22570

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004307 on 310 degrees of freedom
Multiple R-squared: 0.1788,   Adjusted R-squared: 0.1496
F-statistic: 6.135 on 11 and 310 DF, p-value: 4.064e-09

<u>Conclusion</u>: Yes, strong evidence for seasonality. A lot of the dummies are significant (the ones in red). The F-stat=6.135 rejects the H_0 that all seasonal coefficients are 0.

c.
res_sf_ar3_seas <- fit_sf_seas$residuals
> acf(res_sf_ar3_seas)



library(forecast)
fit_auto <- auto.arima(res_lv_ar3_seas)
> fit_auto
Series: res_sf_ar3_seas
ARIMA(2,0,2) with zero mean

Coefficients:

	ar1	ar2	ma1	ma2
	0.9903	-0.8376	-1.1248	0.8890
s.e.	0.0631	0.0828	0.0530	0.0646

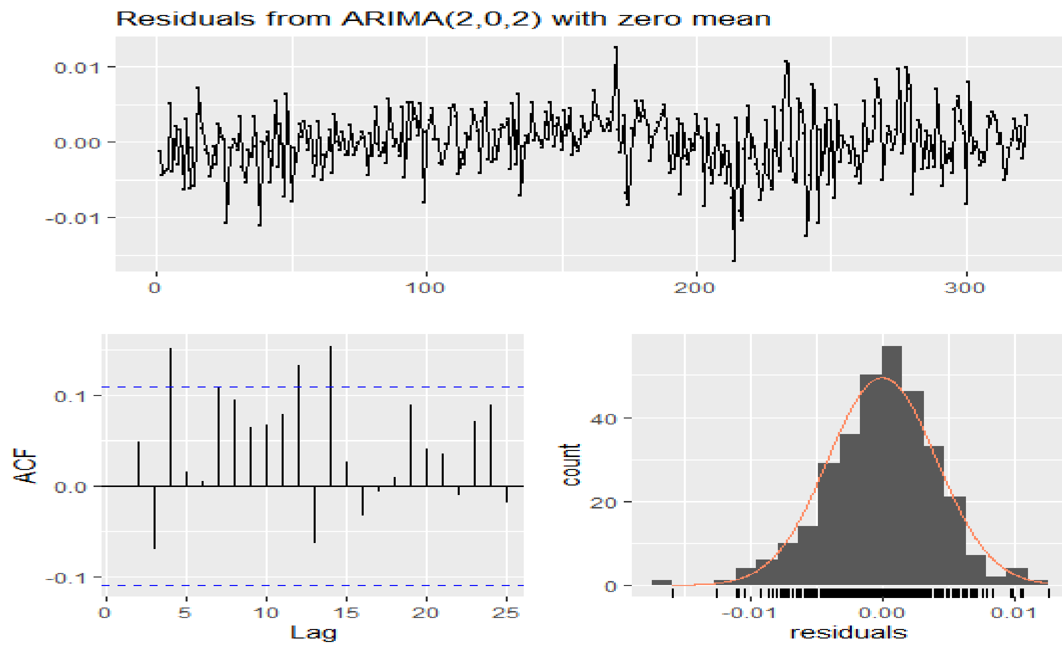
sigma^2 estimated as 1.699e-05: log likelihood=1313.16
AIC=-2616.31   AICc=-2616.12   BIC=-2597.44
> checkresiduals(fit_auto)

Ljung-Box test

data: Residuals from ARIMA(2,0,2) with zero mean
Q* = 19.659, df = 6, p-value = 0.003184

Model df: 4.   Total lags used: 10

> checkresiduals(fit_auto)



Conclusion: Better ACF, but still with significant autocorrelations.

I try an AR(12):

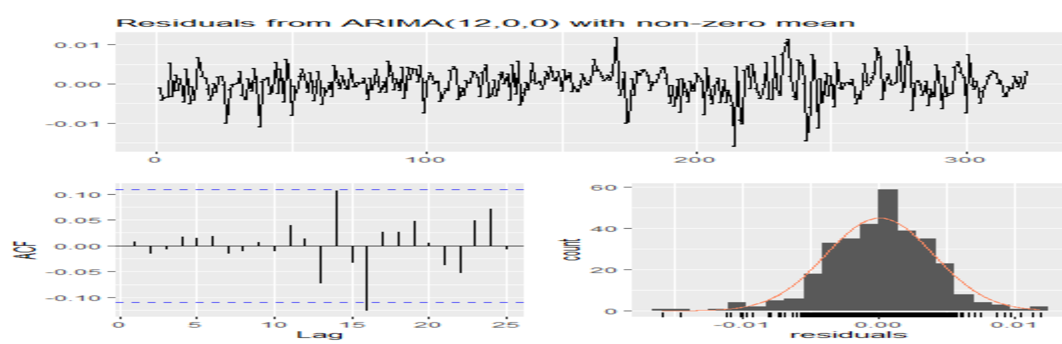
```
fit_ar12 <- arima(res_sf_ar3_seas, order=c(12,0,0))
> fit_ar12

Call:
arima(x = res_sf_ar3_seas, order = c(12, 0, 0))

Coefficients:
          ar1      ar2      ar3     ar4     ar5     ar6     ar7
      -0.1458  -0.0848  -0.1135  0.1975  0.0979  0.0217  0.0820
s.e.   0.0549   0.0551   0.0551  0.0554  0.0565  0.0564  0.0565
          ar8     ar9    ar10    ar11    ar12  intercept
      -0.0042  0.0277  0.0988  0.1136  0.1677      0e+00
s.e.   0.0564  0.0554  0.0551  0.0551  0.0550      4e-04

sigma^2 estimated as 1.553e-05:  log likelihood = 1325.36,  aic = -2622.71

> checkresiduals(fit_ar12)
```



d.

```
f_ar3_coeff <- fit_ar3$coef
T <- length(x_sf1)
y <- x_sf1[4:T]
sf_1 <- x_sf1[3:(T-1)]
sf_2 <- x_sf1[2:(T -2)]
sf_3 <- x_sf1[1:(T-3)]

x0 <- matrix(1,(T-3),1)
xx_sf <- cbind(sf_1, sf_2, sf_3, x0)
y_f0 <- xx_sf[323:nrow(xx_sf),]%*% sf_ar3_coeff
err_f0_ar3 <- y[323:nrow(xx_sf)] - y_f0
mse_ar3 <- sum(err_f0_ar3^2)/length(y_f0)
> mse_ar3
[1] 5.497108e-05
```

1.8

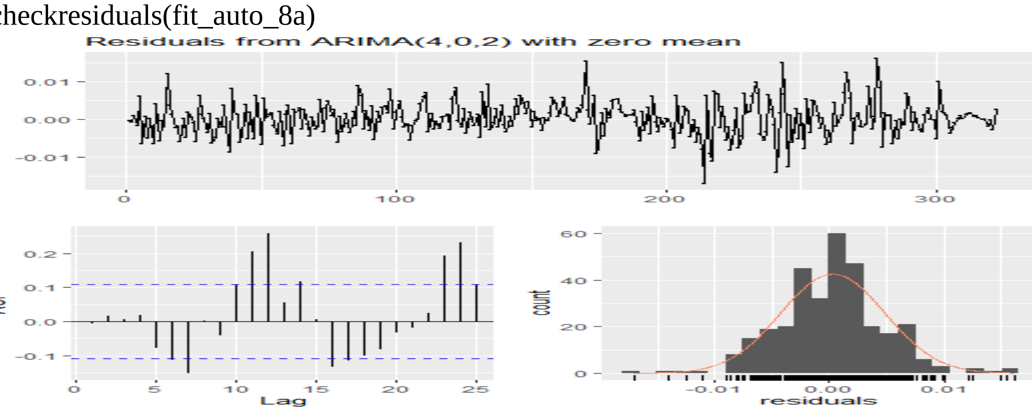
a.

```
library(forecast)
fit_auto_8a <- auto.arima(x_sf)
> fit_auto_8a
Series: x_sf
ARIMA(4,0,2) with zero mean

Coefficients:
         ar1     ar2     ar3      ar4     ma1     ma2
     -0.3853  0.7049  0.5661  -0.2001  1.4618  0.8427
s.e.   0.0948  0.0835  0.0805   0.0651  0.0794  0.0583

sigma^2 estimated as 2.116e-05:  log likelihood=1277.78
AIC=-2541.55   AICc=-2541.2   BIC=-2515.13

> checkresiduals(fit_auto_8a)
```

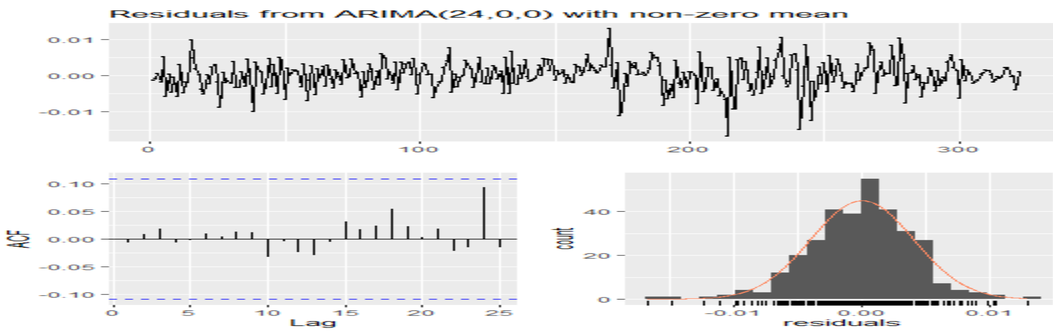


Conclusion: A lot of significant autocorrelations in data. There is a repeated wave pattern.

b

```
fit_arima_8a <- arima(x_sf, order=c(24,0,0))
```

> checkresiduals(fit_arima_8a)



<u>Note</u>: No visible autocorrelation, but a lot of parameters in the model (25). In general, large models are not very efficient. A SARIMA model may work better.

c. Forecast SF_c for the period 2017:1 to 2019:9.

```
sf_ar24_coeff <- fit_arima_8a$coef

lag <- 24
x0_lag <- matrix(1,(T-lag),1)
xx_sf_lag <- cbind(sf_1, sf_2, sf_3, x0_lag)
sf_lag <- matrix(0, (T-lag),lag)
j <- 1
while (j <= lag) {                              # loop to create 24 lags
sf_lag[,j] <- x_sf1[j:(T-lag+j-1)]
j <- j + 1
}

y_f0_arlag <- xx_sf_lag[302:nrow(xx_sf_lag),]%*% sf_ar24_coeff
err_f0_arlag <- y[302:nrow(xx_sf_lag)] - y_f0_arlag
mse_arlag <- sum(err_f0_arlag^2)/length(y_f0_arlag)
> mse_arlag
[1] 0.0001270861                => Higher MSE
```

d.

```
> dm.test(err_f0_arlag,err_f0_ar3, power=2)

        Diebold-Mariano Test

data:  err_f0_arlagerr_f0_ar3
DM = 2.6119, Forecast horizon = 1, Loss function power = 2, p-value = 0.01393
alternative hypothesis: two.sided
```

<u>Conclusion</u>: AR(3) has a better MSE.