Lab 5

PSTAT W 174/274

Seasonal ARIMA models

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as follows:

$$\text{SARIMA: (p, d, q)} \times \text{(P, D, Q)}_s$$

where (p, d, q) is the non-seasonal part of the model and (P, D, Q) is the seasonal part of the model.

$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t$ with $Z_t \sim$ WN(0,${\sigma_Z}^2$) and Y_t=$(1-B)^d(1-B^s)^D X_t$, and $(1-B)^d(1-B^s)^D$ represents the differencing on the original data.

Example 1

Consider a SARIMA: (2, 0, 1) $\times$ (1, 1, 1)$_6$

a. Write the model's equation

- p = 2 then: $\phi(B) = (1 - \phi_1 B - \phi_2 B^2)$
- d = 0 then: $(1 - B)^0 = 1$
- q = 1 then: $\theta(B) = (1 + \theta_1 B)$
- s = 6
- P = 1 then: $\Phi(B) = (1 - \Phi_1 B^6)$
- D = 1 then: $(1 - B^s)^D = (1 - B^6)^1$
- Q = 1 then: $\Theta(B) = (1 + \Theta_1 B^6)$

Finally, we write: $\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t$ as:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^6)(1 - B^6)^1 X_t = (1 + \theta_1 B)(1 + \Theta_1 B^6)Z_t$$

b. How many parameters do you need to estimate for this model?

For the AR components: ϕ_1, ϕ_2, Φ_1, three; For the MA components: θ_1, Θ_1, two; And the white noise variance ${\sigma_Z}^2$, one. In total: 6.

c. Let $Y_t := (1 - B^6)X_t$ and suppose you would like to fit an ARMA(8,7) to Yt. How many parameters would you have to estimate?

In this case we need to estimate 8 terms for the AR component, 7 terms for the MA component and the white noise variance. In total: 16.

d. Why to use SARIMA?

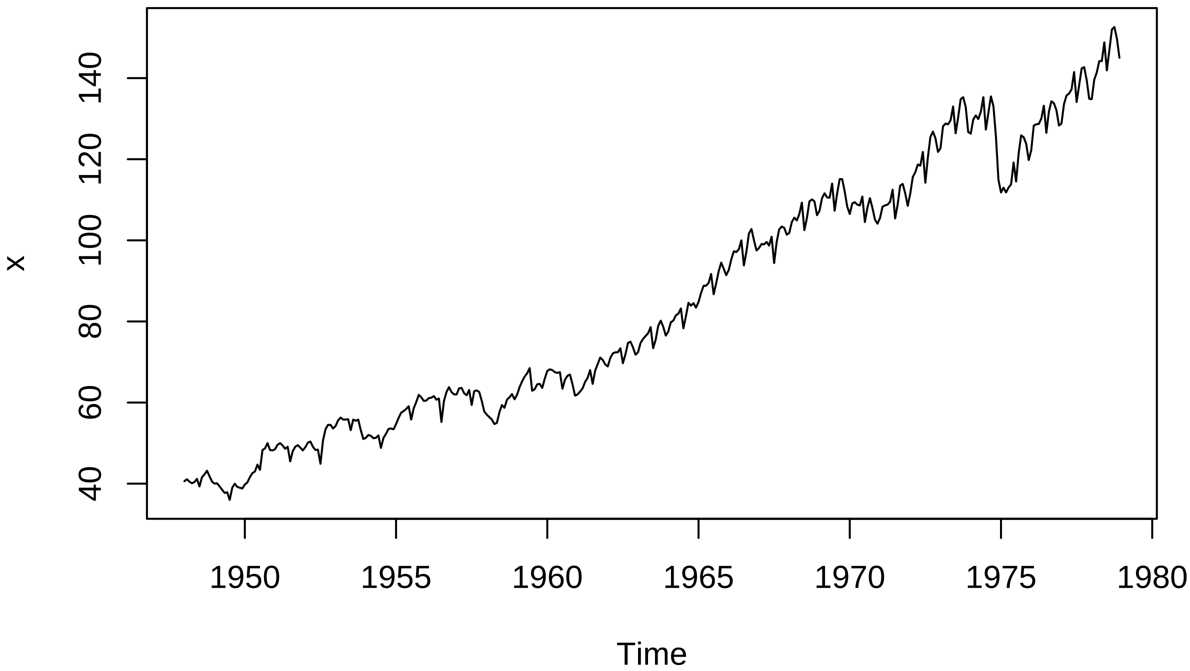
SARIMA is with 6 parameters for estimation from (b) rather than 16 from ARIMA(8,7) in (c).

Example 2: Federal Reserve Board Index

In this case we work with the Monthly Federal Reserve Board Production Index (1948-1978, n = 372 months).

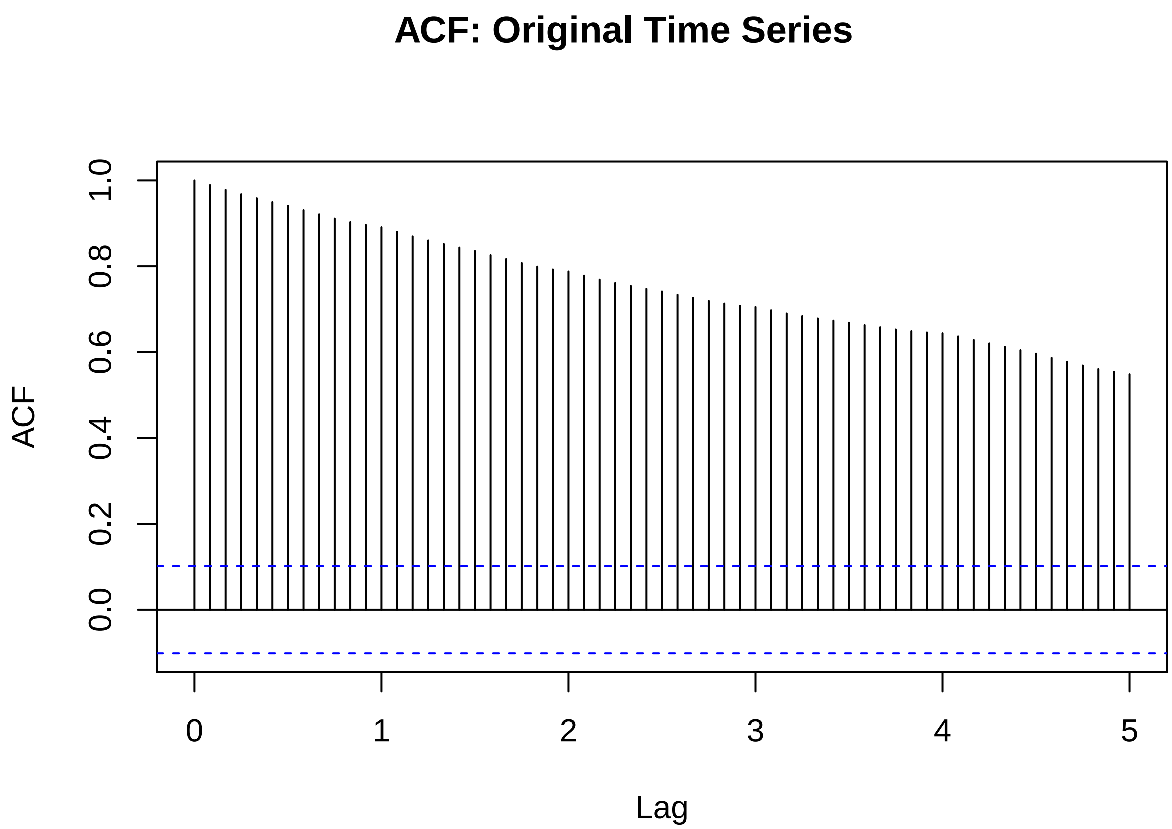
a. Get the production time series and plot it.

```
library(astsa)
x = prodn
plot(x)
```



b. ACF plot.

```
acf(x,lag.max = 60,main = "")
title("ACF: Original Time Series", line = -1, outer=TRUE)
```

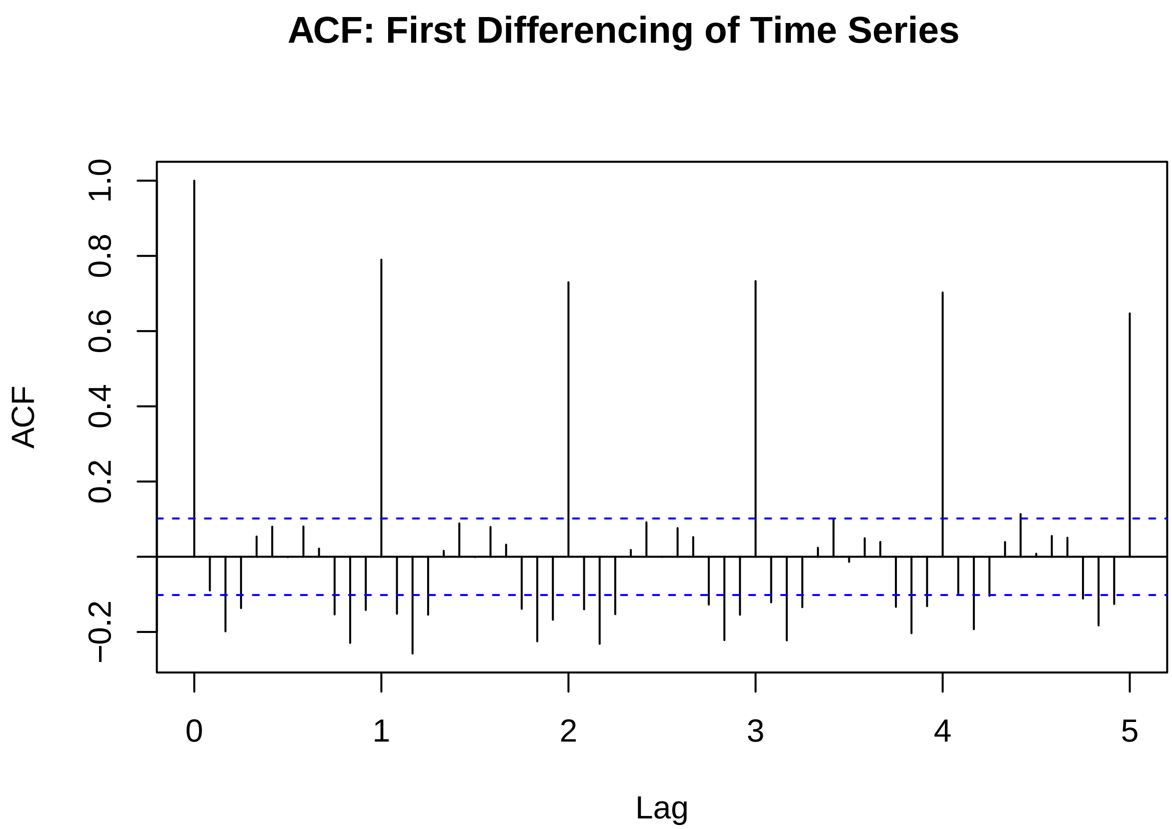


From the plot of the prodn and its acf it is impossible to see whether it is seasonal or not, but there is clear non-stationarity and trend. Therefore, we take difference at lag 1 to remove the trend and produce y_1.

c. Apply a first differencing and plot ACF and PACF.

We plot ACF and PACF plots for $Y_t = \nabla X_t = (1 - B)X_t$, hoping to remove the time trend with the first differencing.

```
y_1 = diff(x, 1)
acf( y_1, lag.max = 60, main = "")
title("ACF: First Differencing of Time Series", line = -1, outer = TRUE)
```



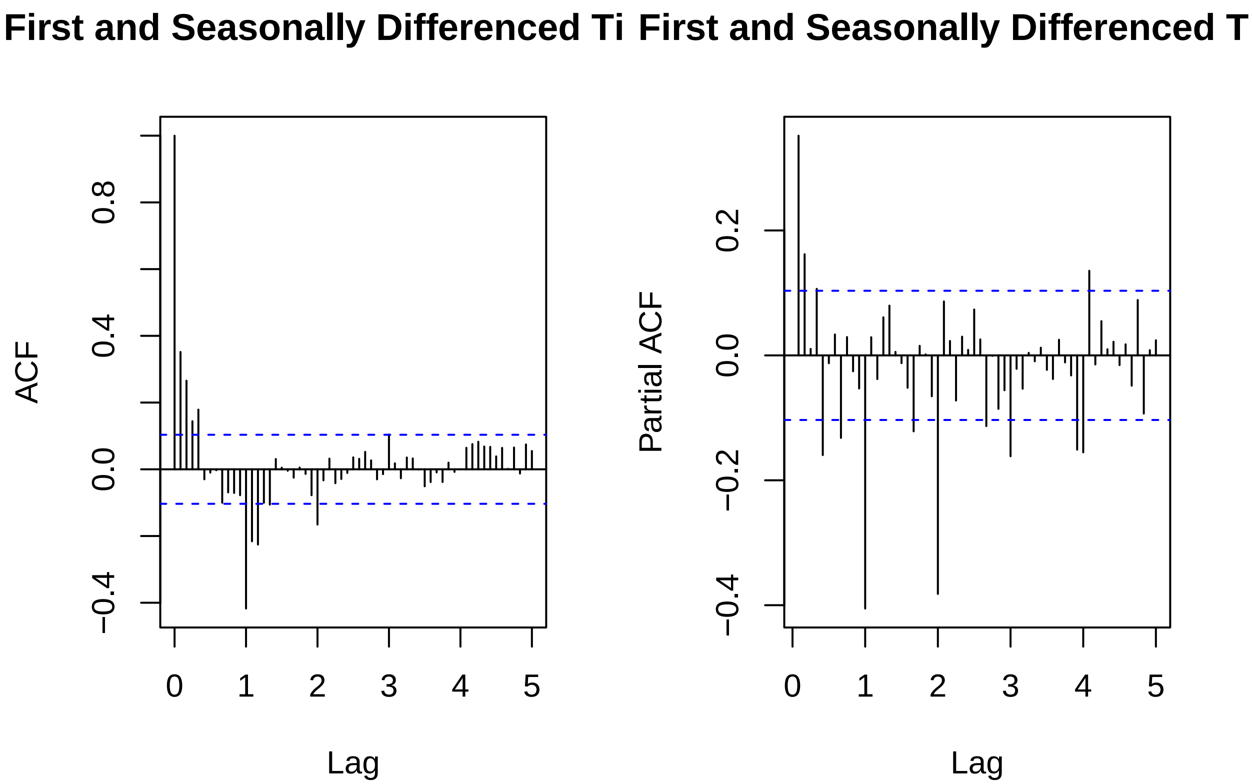
ACF of y_1 shows that the data is nonstationary seasonal (ACF is large even at lag 60). The ACF values are significant at 12∗k, k = 1,2,3,4 and 5. Then in the next step, we second difference at lag 12 to remove seasonal components.

d. Apply a first seasonal differencing and plot ACF and PACF.

In this case, we work with $Y_t = \nabla_{12}\nabla X_t = (1 - B^{12})(1 - B)X_t$

```
op <- par(mfrow = c(1,2))
y_12 = diff(y_1, 12)
acf( y_12, lag.max = 60, main = "ACF: First and Seasonally Differenced Time Series")

pacf( y_12, lag.max = 60, main = "PACF: First and Seasonally Differenced Time Series")
```



e. Based on part d), suggest some models to fit.

Modeling the seasonal part (P, D, Q): For this part, focus on the seasonal lags h = 1s, 2s, etc.

- We applied one seasonal differencing so D = 1 at lag s = 12.
- The ACF shows a strong peak at h = 1s and smaller peaks appearing at h = 2s, 3s.

  A good choice for the MA part could be Q=1 or 2 but not 3. This is because acf at lag 36 sits on the border of the confidence interval and using Bartlett's formula, one knows that these borders are too conservative for MA models.
- The PACF shows two strong peaks at h = 1s, 2s and smaller peaks at h = 3s, 4s.

  A good choice for the AR part could be P = 2 or P = 4.

Modeling the non-seasonal part (p , d, q): In this case focus on the within season lags, h = 1,. . . ,11.

- We applied one differencing to remove the trend: d = 1
- The ACF seems to be tailing off. Or perhaps cuts off at lag.

  good choice for the MA part could be q = 0 or q = 4 respectively.
- The PACF cuts off at lag h=2 or 5.

  A good choice for the AR part could be p = 2 or p = 5.

f. Fit a couple of candidate models

As an illustration we fit the following two models:

i. SARMA(p=2,d=1,q=0)×(P =2,D=1,Q=1)$_{s=12}$

ii. SARMA(p=2,d=1,q=4)×(P =4,D=1,Q=2)$_{s=12}$

Model 1:

```
fit.i <- sarima( xdata = prodn,p = 2, d = 1, q = 0,
                 P = 2 , D = 1, Q = 2, S = 12, details = F)
fit.i$fit
```

Fitting the model:

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1     ar2    sar1     sar2     sma1    sma2
##       0.3019  0.0945  0.3938  -0.2942  -1.1445  0.4951
## s.e.  0.0528  0.0535  0.1589   0.0751   0.1618  0.1126
##
## sigma^2 estimated as 1.322:  log likelihood = -565.19,  aic = 1144.37
```

Checking coefficients: From the above coefficients table, ar2 is not significant because the confidence interval of the estimated coefficient contains 0. Therefore, we should set these coefficients to 0.

```
fit.i1 <- sarima( xdata = prodn,p = 2, d = 1, q = 0,
                 P = 2 , D = 1, Q = 2, S = 12, details = F, fixed=c(NA,0,NA,NA,NA,NA))
fit.i1$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1  ar2    sar1     sar2     sma1    sma2
##       0.3328    0  0.4004  -0.2866  -1.1678  0.5098
## s.e.  0.0501    0  0.1510   0.0751   0.1522  0.1071
##
## sigma^2 estimated as 1.332:  log likelihood = -566.74,  aic = 1145.48
```

From the above output, we can write the model as:

$$(1 - 0.3328B)(1 - 0.4004B^{12} + 0.2866B^{24})(1 - B)(1 - B^{12})X_t = (1 - 1.1678B^{12} + 0.5098B^{24})Z_t$$

So, it is a SARMA(p=1,d=1,q=0)×(P =2,D=1,Q=2)$_{s=12}$

Check the model stationarity/invertibility: Lastly, we check the model stationarity/invertibility.

AR part

$$(1 - 0.3328B)$$

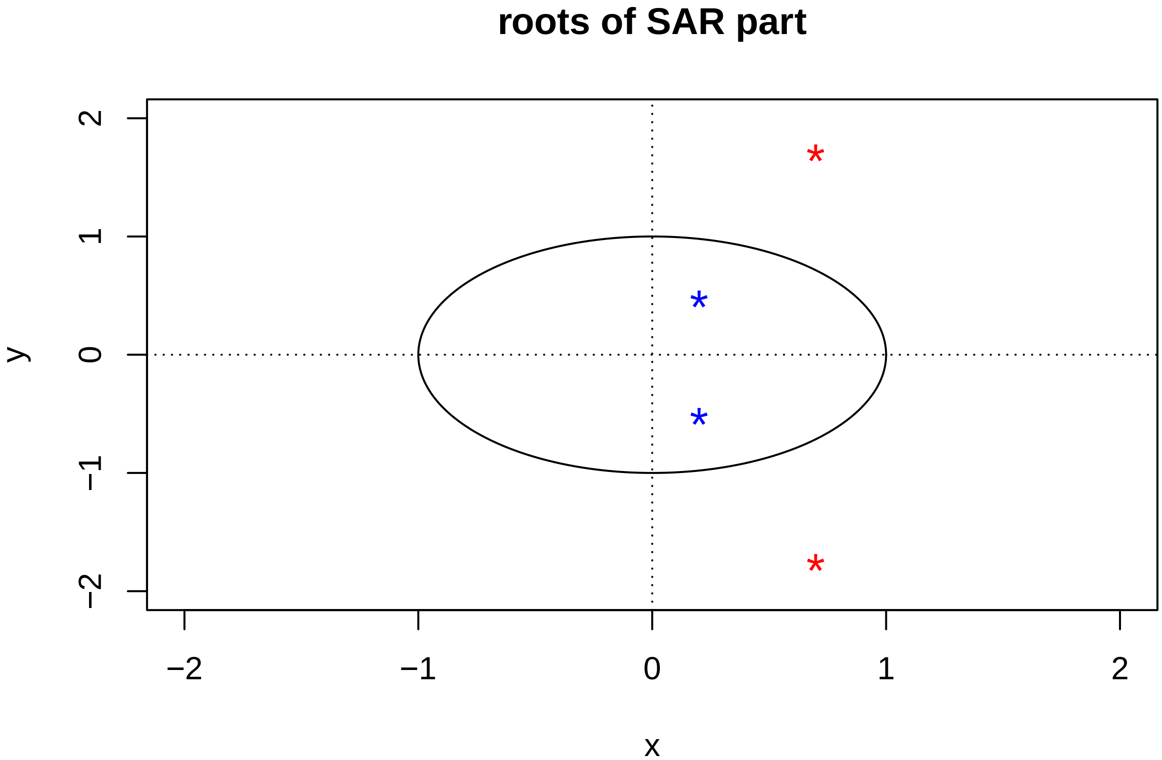
: 0.3328

$$<$$

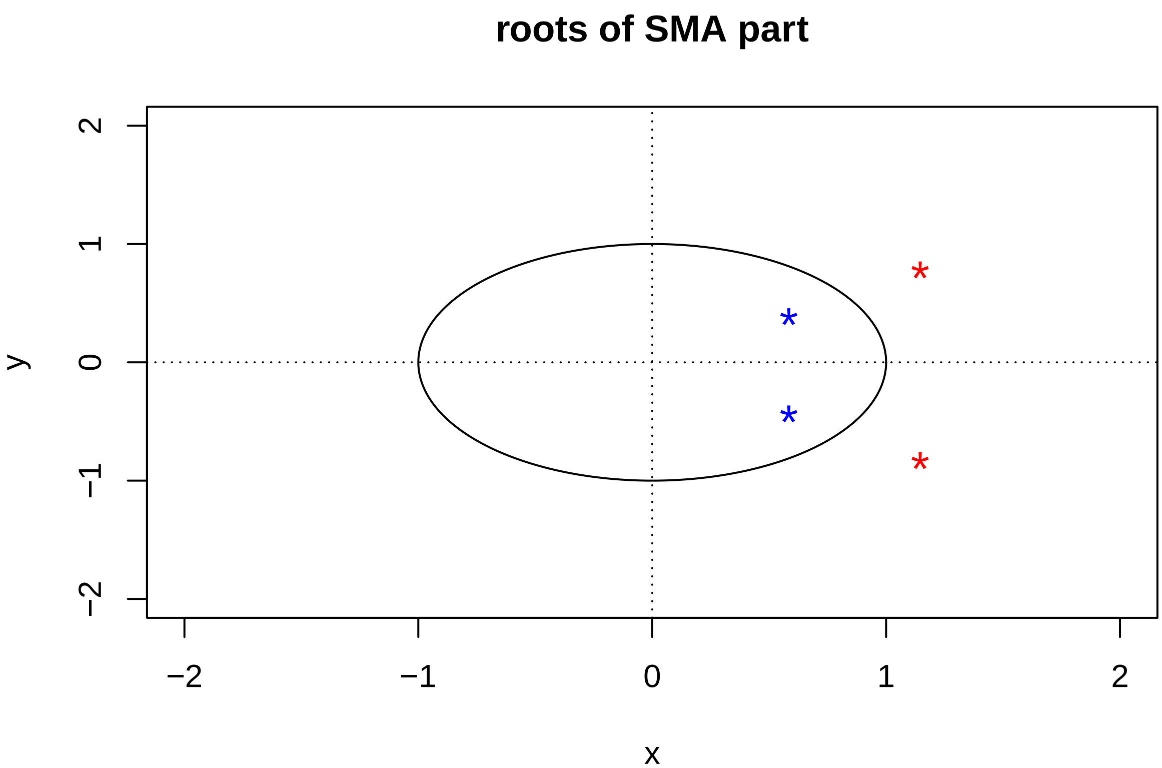
1, so the root is greater than 1. MA part: no MA part in this model

For the SAR and SMA part, we can use the plot.roots.R in week 5 module to plot the roots:

```
plot.roots(NULL,polyroot(c(1, -0.4004,0.2866)), main="roots of SAR part")
```



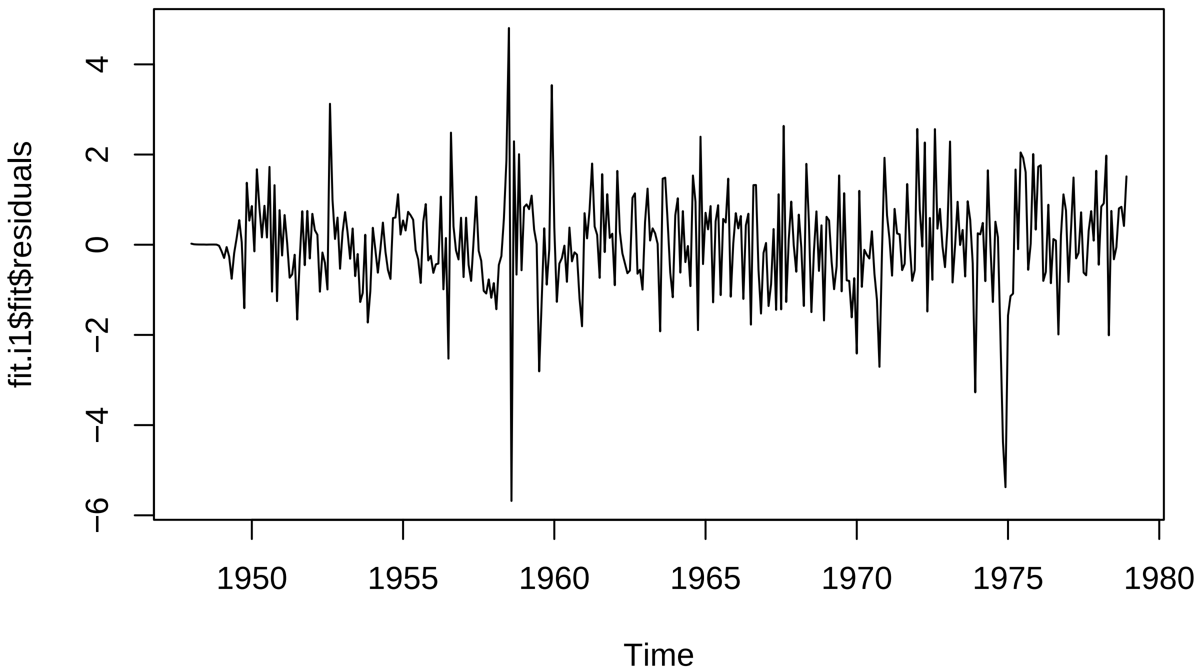

```
plot.roots(NULL,polyroot(c(1, -1.1678,0.5098)), main="roots of SMA part")
```



From the above two plots, the roots are outside the unit circle (blue dots are all within the unit circle). Therefore, the model is stationary and invertible.

Plot the residuals: Here is the residuals of the first model:

```
plot(fit.i1$fit$residuals)
```



Model 2:

Fitting the model: We repeat similar process to our second candidate model:

```
fit.ii <- sarima( xdata = prodn,p = 2, d = 1, q = 4,
                  P = 4 , D = 1, Q = 2, S = 12,
                  details = F)

fit.ii$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##           ar1      ar2     ma1     ma2     ma3     ma4     sar1     sar2
##       -0.6156  -0.3924  0.9286  0.7805  0.2782  0.2243  -0.5480  -0.7162
## s.e.   0.4134   0.3373  0.4077  0.4438  0.1547  0.0589   0.7545   0.2443
##          sar3     sar4     sma1    sma2
##       -0.3389  -0.2605  -0.1705  0.1688
## s.e.   0.2548   0.0856   0.7772  0.3558
##
## sigma^2 estimated as 1.284:  log likelihood = -559.96,  aic = 1145.93
```

Checking coefficients: From the above coefficients table, ar1, ar2, ma2, ma3, sar1, sar3, sma1 and sma2 are not significant because the confidence interval of the estimated coefficient contains 0. Therefore, we should set these coefficients to 0.

```
fit.ii1 <- sarima( xdata = prodn,p = 2, d = 1, q = 4,
                   P = 4 , D = 1, Q = 2, S = 12,
                   details = F,fixed=c(0,0,NA,0,0,NA,0,NA,0,NA,0,0))
fit.ii1$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##       ar1  ar2     ma1  ma2  ma3     ma4  sar1     sar2  sar3     sar4  sma1
##         0    0  0.2768    0    0  0.2195     0  -0.1812     0  -0.0711     0
## s.e.    0    0  0.0425    0    0  0.0444     0   0.0532     0   0.0578     0
##       sma2
##          0
## s.e.     0
##
## sigma^2 estimated as 2.11:  log likelihood = -644.01,  aic = 1298.02
```

Then from the above output of the updated model, we find out that sar4 turns out not significant. Therefore, we fix the coefficient of sar4 to be 0 and fit the model one more time:

```
fit.ii2 <- sarima( xdata = prodn,p = 2, d = 1, q = 4,
                   P = 4 , D = 1, Q = 2, S = 12,
                   details = F,fixed=c(0,0,NA,0,0,NA,0,NA,0,0,0,0))
fit.ii2$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##       ar1  ar2     ma1  ma2  ma3     ma4  sar1     sar2  sar3  sar4  sma1  sma2
##         0    0  0.2748    0    0  0.2111     0  -0.1684     0     0     0     0
## s.e.    0    0  0.0428    0    0  0.0437     0   0.0521     0     0     0     0
##
## sigma^2 estimated as 2.12:  log likelihood = -644.76,  aic = 1297.53
```

From the above output, we can write the model as:

$$(1+0.1684B^{24})(1-B)(1-B^{12})X_t = (1+0.2748B+0.2111B^4)Z_t$$

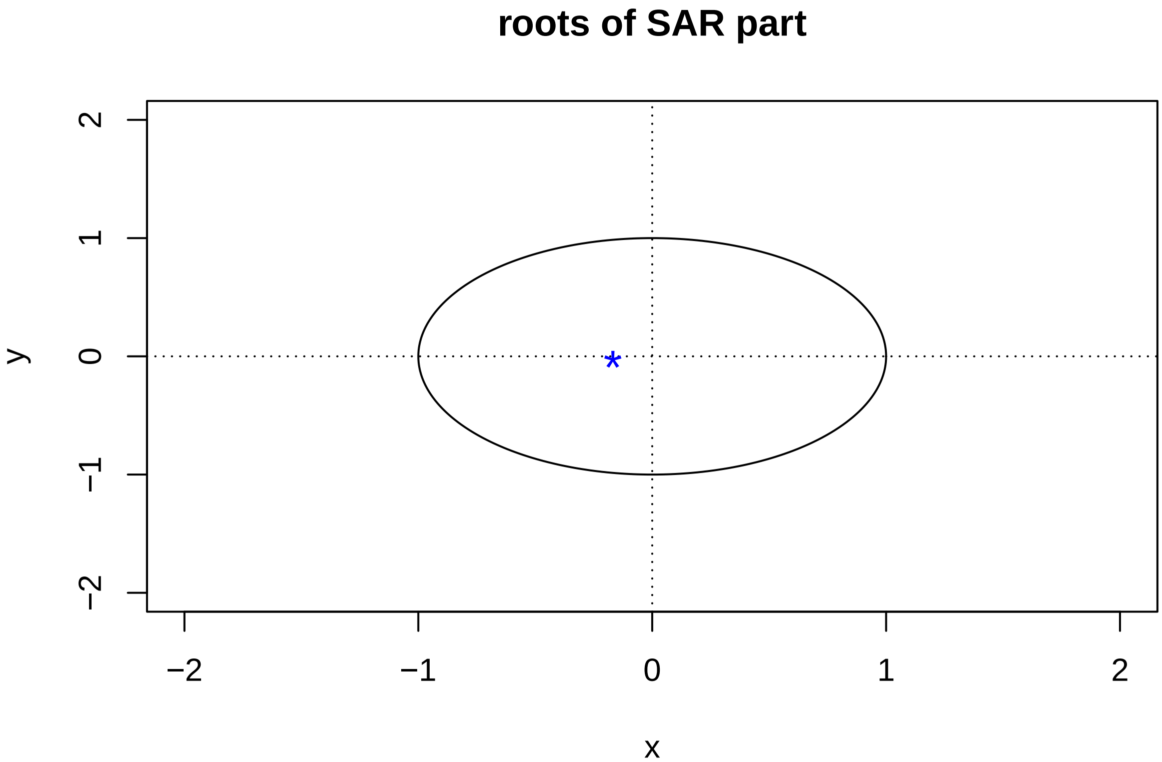
Check the model stationarity/invertibility:

Lastly, we check the model stationarity/invertibility.

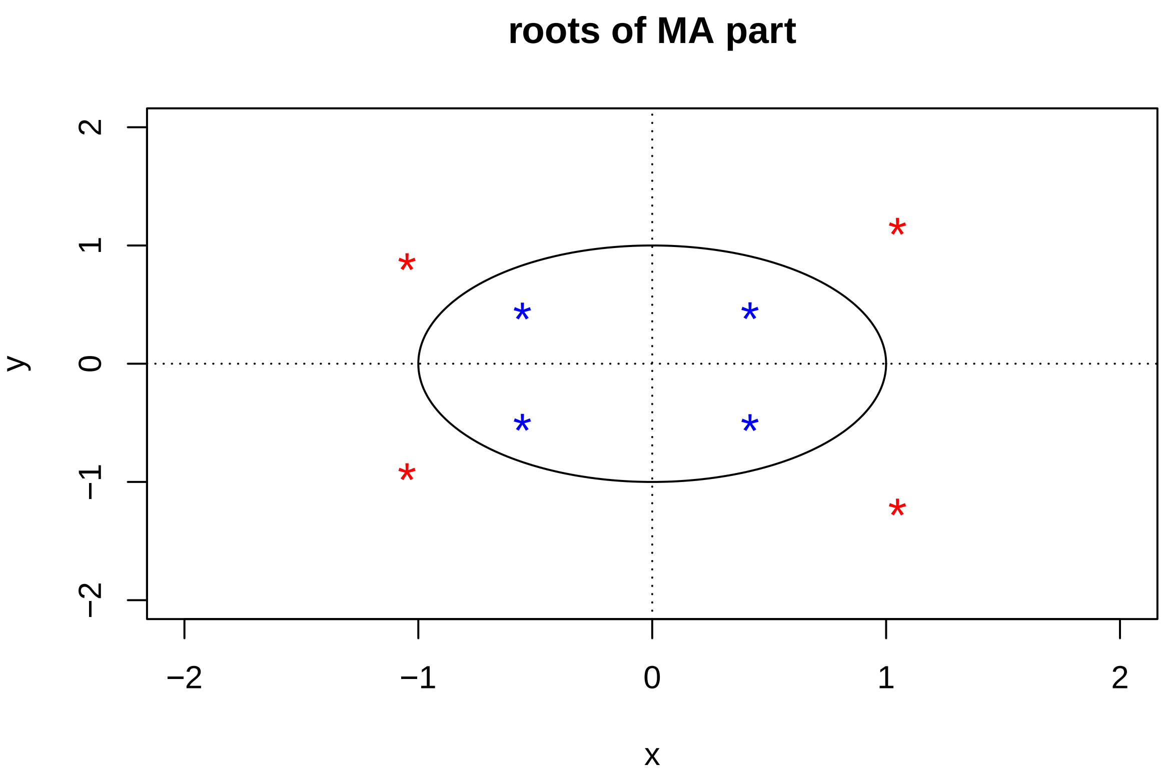
AR part: no AR part in this model. SMA part: no SMA part in this model.

For the SAR and MA part, we can use the plot.roots.R in week 5 module to plot the roots:

```
plot.roots(NULL,polyroot(c(1, 0.1684)), main="roots of SAR part")
```




```
plot.roots(NULL,polyroot(c(1, 0.2748,0, 0, 0.2111)), main="roots of MA part")
```



From the above two plots, the roots are outside the unit circle (blue dots are all within the unit circle).

Therefore, the model is stationary and invertible.

Plot the residuals: Here is the residuals of the second model:

```
plot(fit.ii1$fit$residuals)
```