STA457TUT7

Shu Wang

2022-03-30

```
install.packages('astsa')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library('astsa')
data(globtemp)
```

(a)

A process x_t is said to be ARIMA(p,d,q) if:

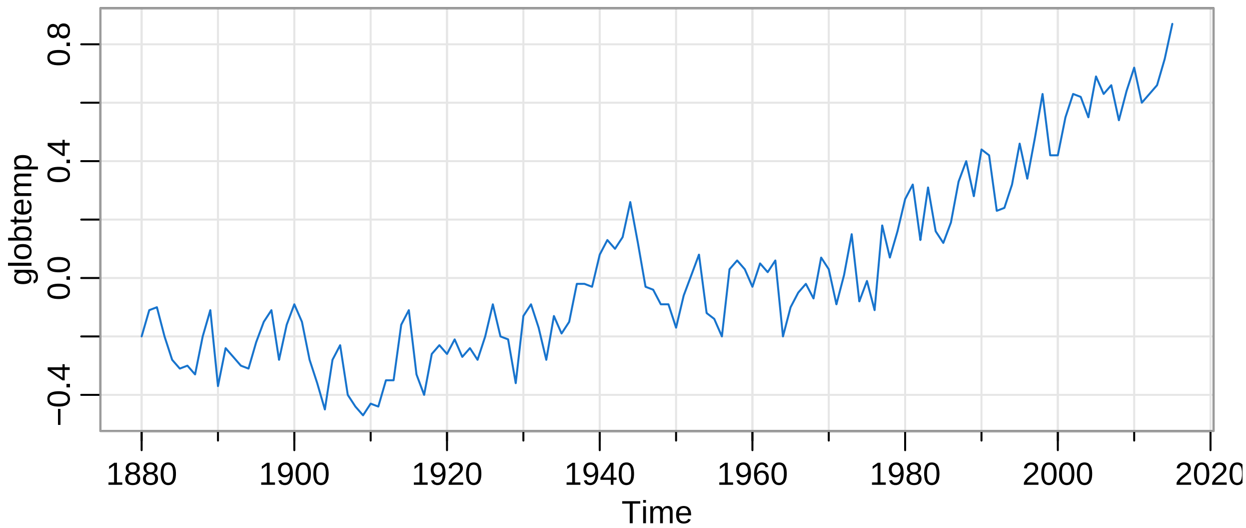
$\bigtriangledown^k x_t = (1-B)^d x_t$

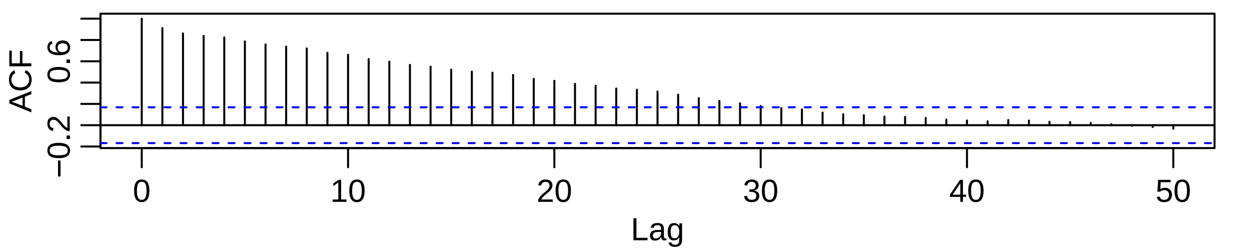
is ARMA(p,q).

The model is written as: $\phi(B)(1-B)^d x_t = \alpha + \theta(B)\omega_t$ where $\alpha = \delta(1-\phi_1 - ... - \phi_p)$ and $\delta = E(\bigtriangledown^k x_t)$

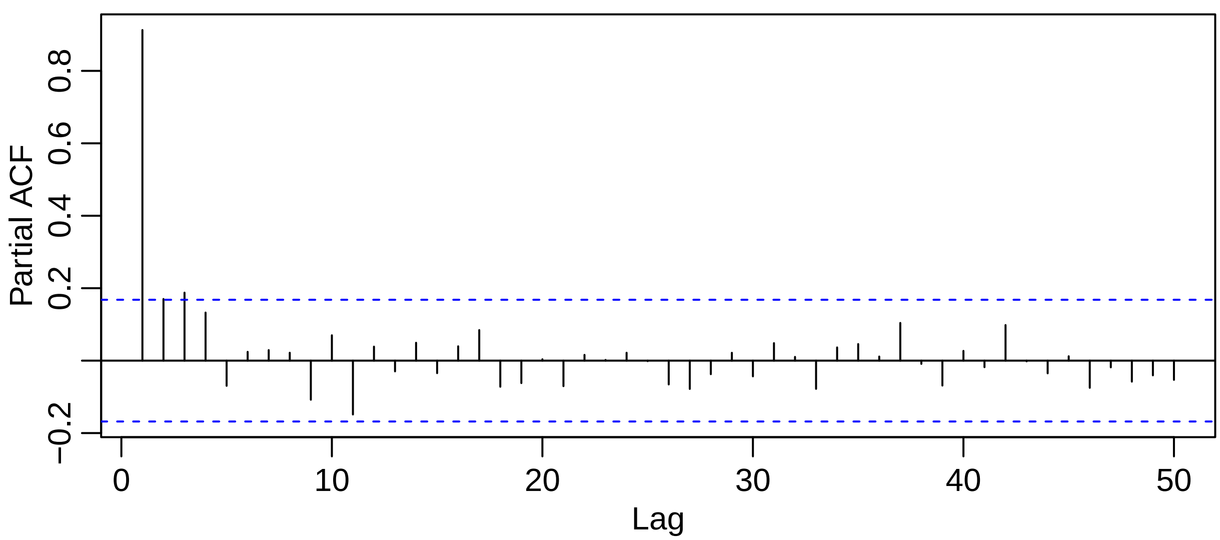
After differencing,AR and MA dependence structures may exist: ARIMA(p, d ,q) – p : AR(p) – value at time t depends on previous p values) – d : # of differences (need to take d th difference to make stationary) – q : MA(q) – value at time t depends on previous q random shocks)

```
layout(1:2,heights = 2:1)
tsplot(globtemp, col = 4)
acf(globtemp,lag.max = 50)
```



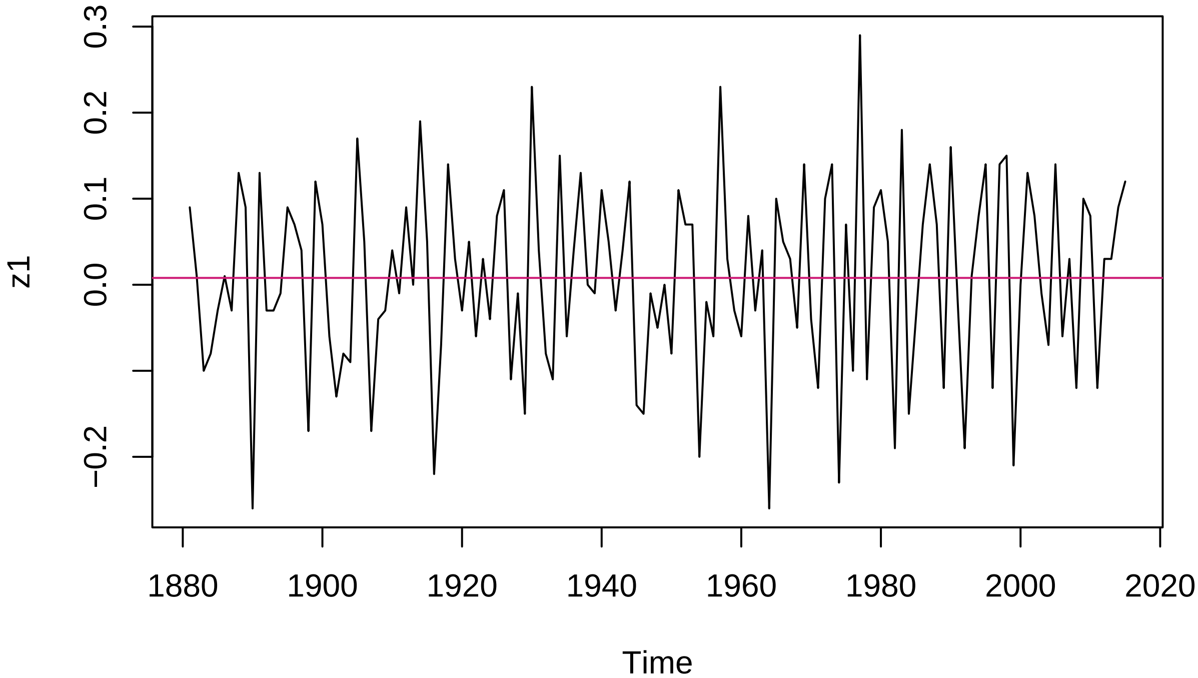



```
pacf(globtemp,lag.max = 50)
```

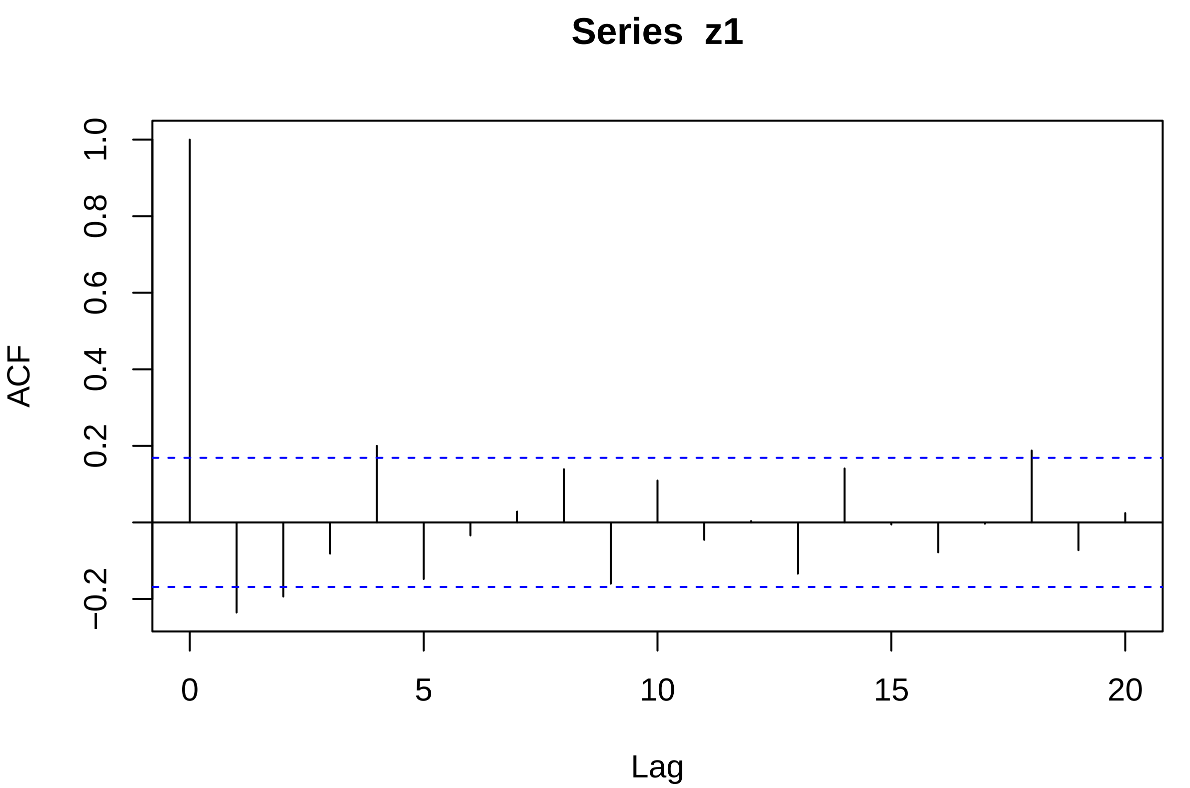


Since many of the values are below 0,it is not appropriate to perform a log transformation on the data.Instead,we can do differencing on the data.

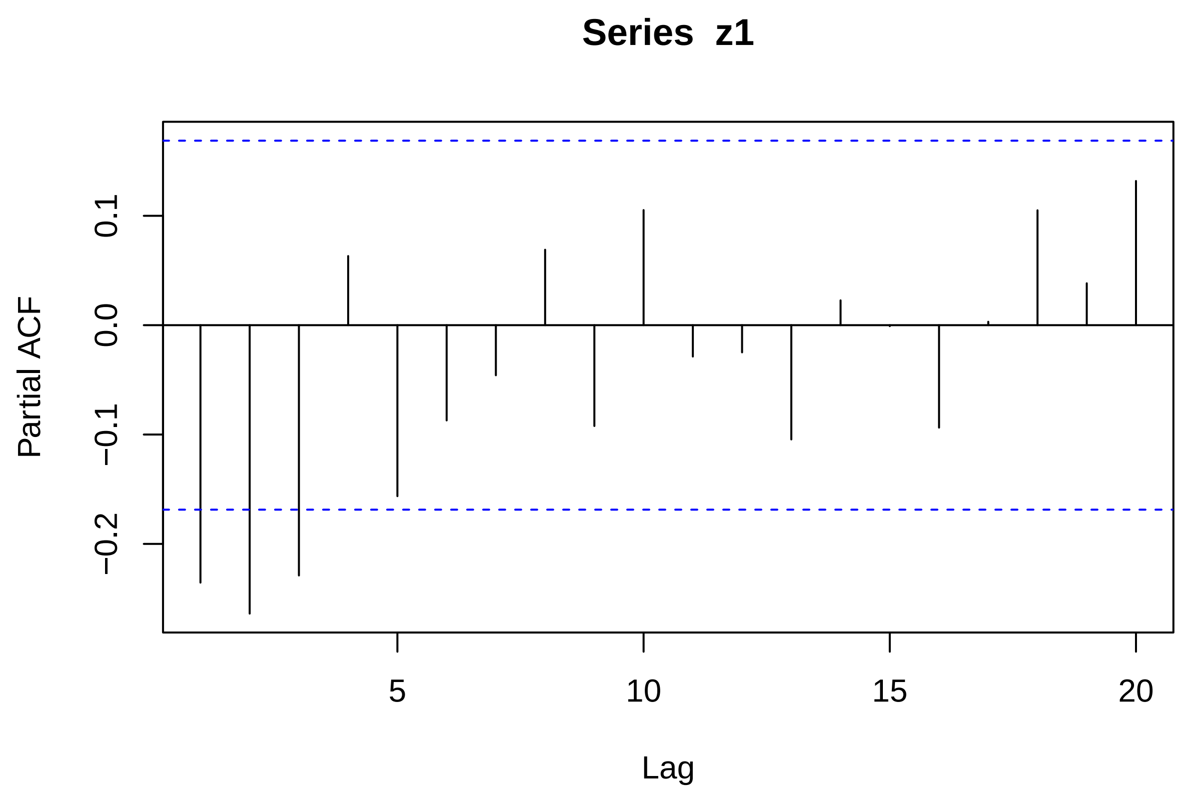
```
z1 = diff(globtemp)
plot(z1)
abline(h = mean(z1),col = 6)
```




```
acf(z1, 20)
```




```
pacf(z1,20)
```




```r
install.packages('lmtest')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```r
library('lmtest')
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

We perform a augmented dickey-fuller test to test the stationary of the first-order differencing time series.

```r
install.packages('tseries')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```r
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
#perform augmented Dickey-Fuller test
adf.test(z1)
```

```
## Warning in adf.test(z1): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  z1
## Dickey-Fuller = -6.79, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

We have strong evidence against the null hypothesis that the process is non-stationary.

We can see that ACF is cutting off at lag 2 and the PACF is tailing off. This would suggest the globtemp follows an MA(2) process, or diff(globtemp) follows an ARIMA(0, 1, 2) model.

Rather than focus on one model, we will also suggest that it appears that the ACF is tailing off and the PACF is cutting off at lag 3.

This suggests an AR(3) model for the globtemp, or ARIMA(3, 1, 0) for z1.

(b)

```r
install.packages("forecast")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

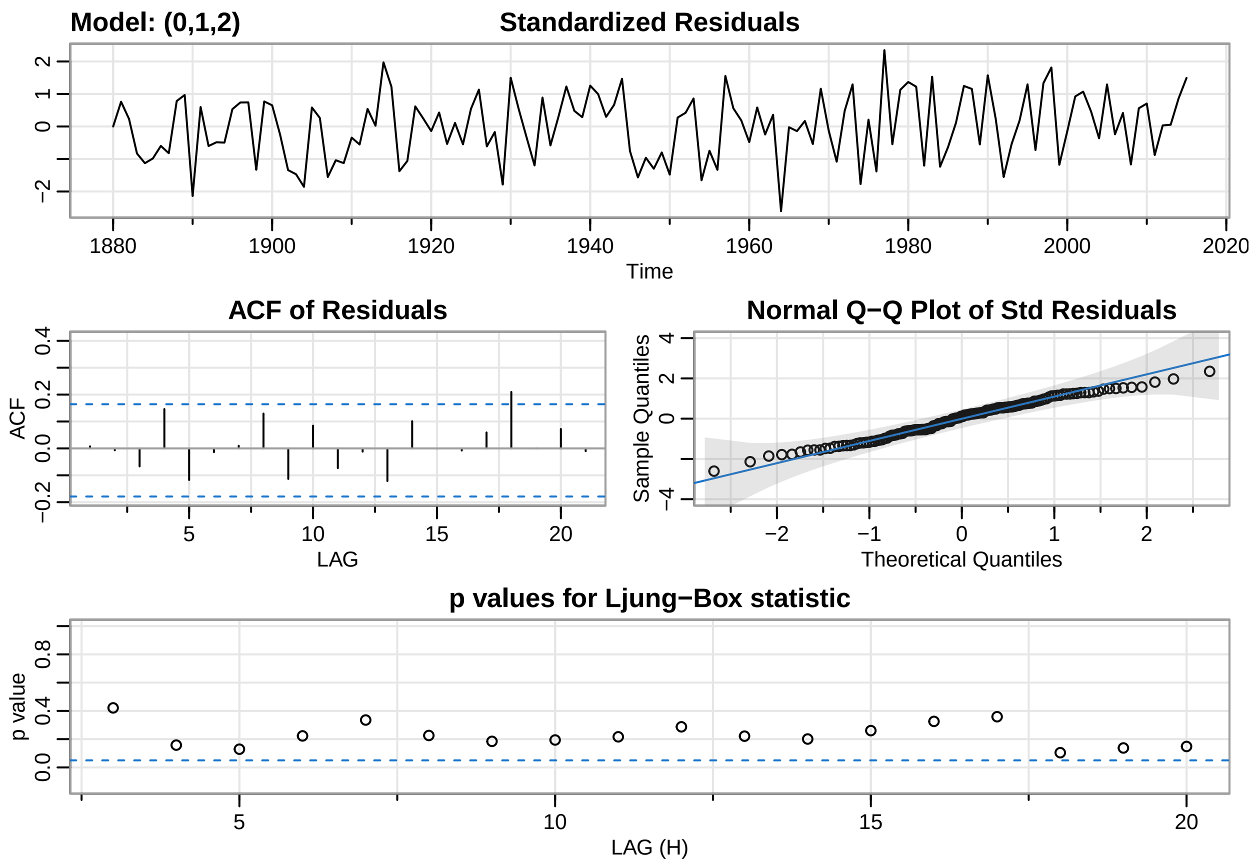
```r
library(forecast)
```

```
##
## Attaching package: 'forecast'

## The following object is masked from 'package:astsa':
##
##     gas
```

```r
sarima(globtemp,p=0,d=1,q=2)
```

```
## initial  value -2.220513
## iter   2 value -2.294887
## iter   3 value -2.307682
## iter   4 value -2.309170
## iter   5 value -2.310360
## iter   6 value -2.311251
## iter   7 value -2.311636
## iter   8 value -2.311648
## iter   9 value -2.311649
## iter   9 value -2.311649
## iter   9 value -2.311649
## final  value -2.311649
## converged
## initial  value -2.310187
## iter   2 value -2.310197
## iter   3 value -2.310199
## iter   4 value -2.310201
## iter   5 value -2.310202
## iter   5 value -2.310202
## iter   5 value -2.310202
## final  value -2.310202
## converged
```

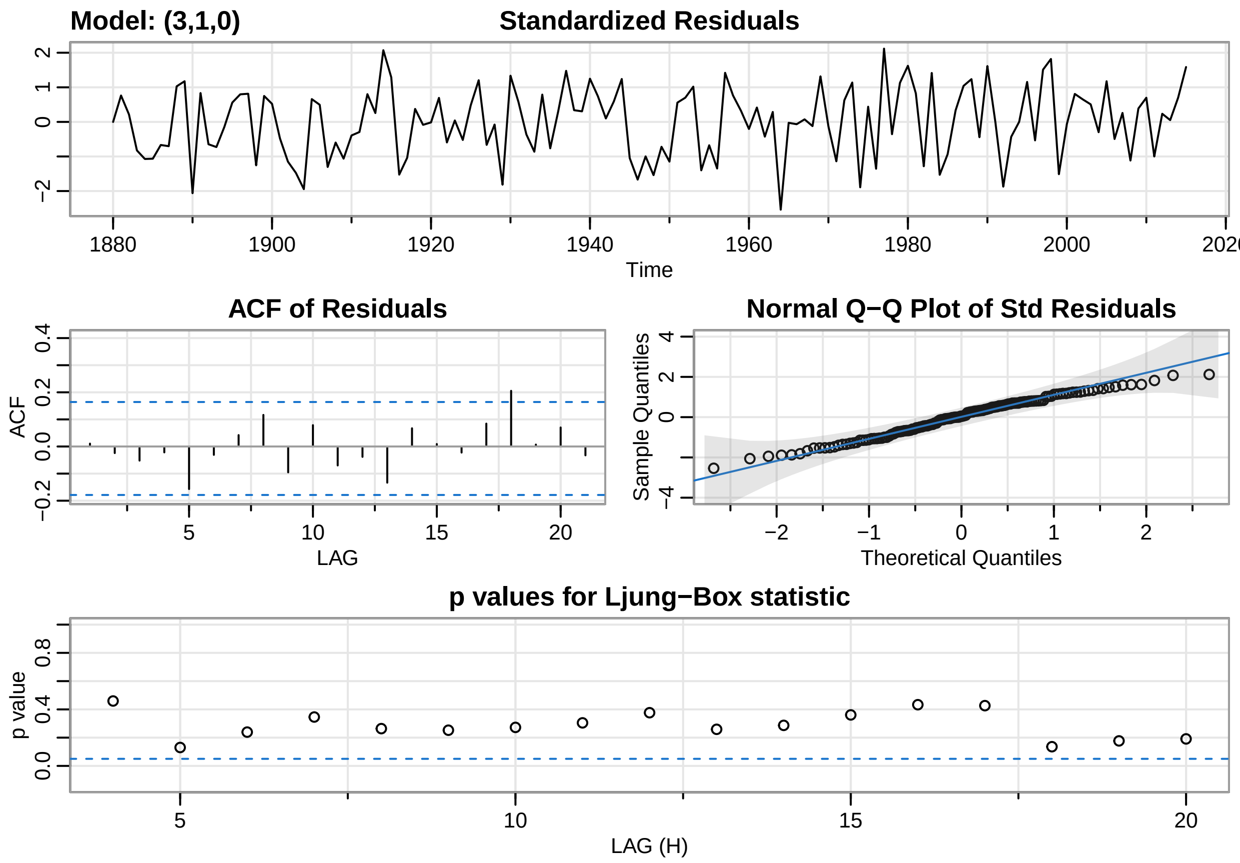



```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     xreg = constant, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##           ma1      ma2  constant
##       -0.3984  -0.2173    0.0072
## s.e.   0.0808   0.0768    0.0033
##
## sigma^2 estimated as 0.00982:  log likelihood = 120.32,  aic = -232.64
##
## $degrees_of_freedom
## [1] 132
##
## $ttable
##          Estimate     SE t.value p.value
## ma1       -0.3984 0.0808 -4.9313  0.0000
## ma2       -0.2173 0.0768 -2.8303  0.0054
## constant   0.0072 0.0033  2.1463  0.0337
##
## $AIC
## [1] -1.723268
##
```

```
## $AICc
## [1] -1.721911
##
## $BIC
## [1] -1.637185
```

```
sarima(globtemp,p=3,d=1,q=0)
```

```
## initial  value -2.215090
## iter   2 value -2.289035
## iter   3 value -2.306884
## iter   4 value -2.308838
## iter   5 value -2.309367
## iter   6 value -2.309746
## iter   7 value -2.309749
## iter   7 value -2.309749
## iter   7 value -2.309749
## final  value -2.309749
## converged
## initial  value -2.314672
## iter   2 value -2.314677
## iter   3 value -2.314679
## iter   4 value -2.314682
## iter   4 value -2.314682
## iter   4 value -2.314682
## final  value -2.314682
## converged
```



```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     xreg = constant, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##           ar1      ar2      ar3  constant
##       -0.3669  -0.3421  -0.2363    0.0071
## s.e.   0.0842   0.0849   0.0838    0.0044
##
## sigma^2 estimated as 0.009733:  log likelihood = 120.93,  aic = -231.85
##
## $degrees_of_freedom
## [1] 131
##
## $ttable
##          Estimate     SE t.value p.value
## ar1       -0.3669 0.0842 -4.3592  0.0000
## ar2       -0.3421 0.0849 -4.0319  0.0001
## ar3       -0.2363 0.0838 -2.8186  0.0056
## constant   0.0071 0.0044  1.6168  0.1083
##
## $AIC
## [1] -1.717413
##
## $AICc
## [1] -1.715133
##
## $BIC
## [1] -1.60981
```

We now have two models,with ARIMA(0,1,2) and ARIMA(3,1,0) separately.

For the ARIMA(0,1,2) model,the p_value for all MA parameters are <0.05. At 5% significance level,we have strong evidence against the null hypothesis that the parameter θ_1,$theta_2$ and the constant term is 0.

For the ARIMA(3,1,0) model,the p_value for all AR parameters(except the constant term) are <0.05.At 5% significance level,we have strong evidence against the null hypothesis that the parameter ϕ_1,phi_2,ϕ_3 are 0. The p_value for constant term is 0.10. At 5% significance level,we can conclude that the constant term is 0.

(c)

Residual diagnostic check:

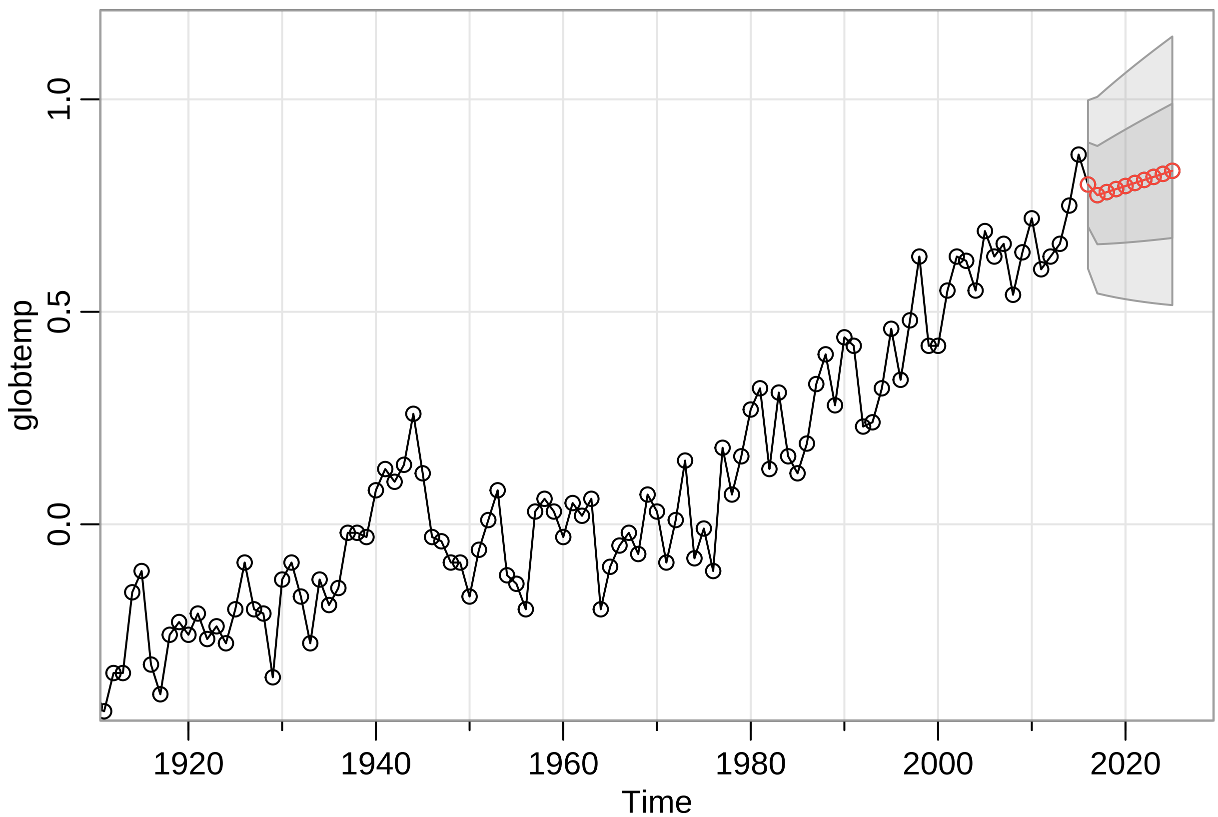
If the model fits well, the standardized residuals should behave as an iid sequence with mean zero and variance one.As we can see from sarima results in part(b),the residuals appears to be i.i.d with mean 0 and variance 1,the QQ-plot also indicates the normality of the residuals are hold.The Q test also shows that there's no apparent autocorrelation left in the residuals for both models.

We reserve the constant in the ARIMA(0,1,2) model,since the constant is significant.

(d)

To choose the final model, we compare the AIC, the AICc, and the BIC for both models. The AIC,AICc and BIC for ARIMA(0,1,2) is lower,which suggests we prefer the ARIMA(0,1,2) model.

```
prediction<-sarima.for(globtemp, n.ahead = 10, p=0, d=1, q=2)
```




```
predict = as.numeric(prediction$pred)
CI_lower = as.numeric(prediction$pred - qnorm(0.95) * prediction$se)
CI_upper = as.numeric(prediction$pred + qnorm(0.95) * prediction$se)
Year = c("2016","2017","2018","2019","2020","2021","2022","2023","2024","2025")
my_data <- data.frame(Year,predict,CI_lower,CI_upper)
my_data
```

```
##    Year   predict  CI_lower  CI_upper
## 1  2016 0.7995567 0.6365590 0.9625544
## 2  2017 0.7745381 0.5843177 0.9647584
## 3  2018 0.7816919 0.5814214 0.9819624
## 4  2019 0.7888457 0.5790059 0.9986855
## 5  2020 0.7959996 0.5770082 1.0149909
## 6  2021 0.8031534 0.5753779 1.0309289
## 7  2022 0.8103072 0.5740740 1.0465405
## 8  2023 0.8174611 0.5730626 1.0618596
## 9  2024 0.8246149 0.5723153 1.0769146
## 10 2025 0.8317688 0.5718080 1.0917295
```