Exam 3 Code

Trelan Hakeem-Barron

3/27/2020

```
library(TSA)
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(forecast)
```

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

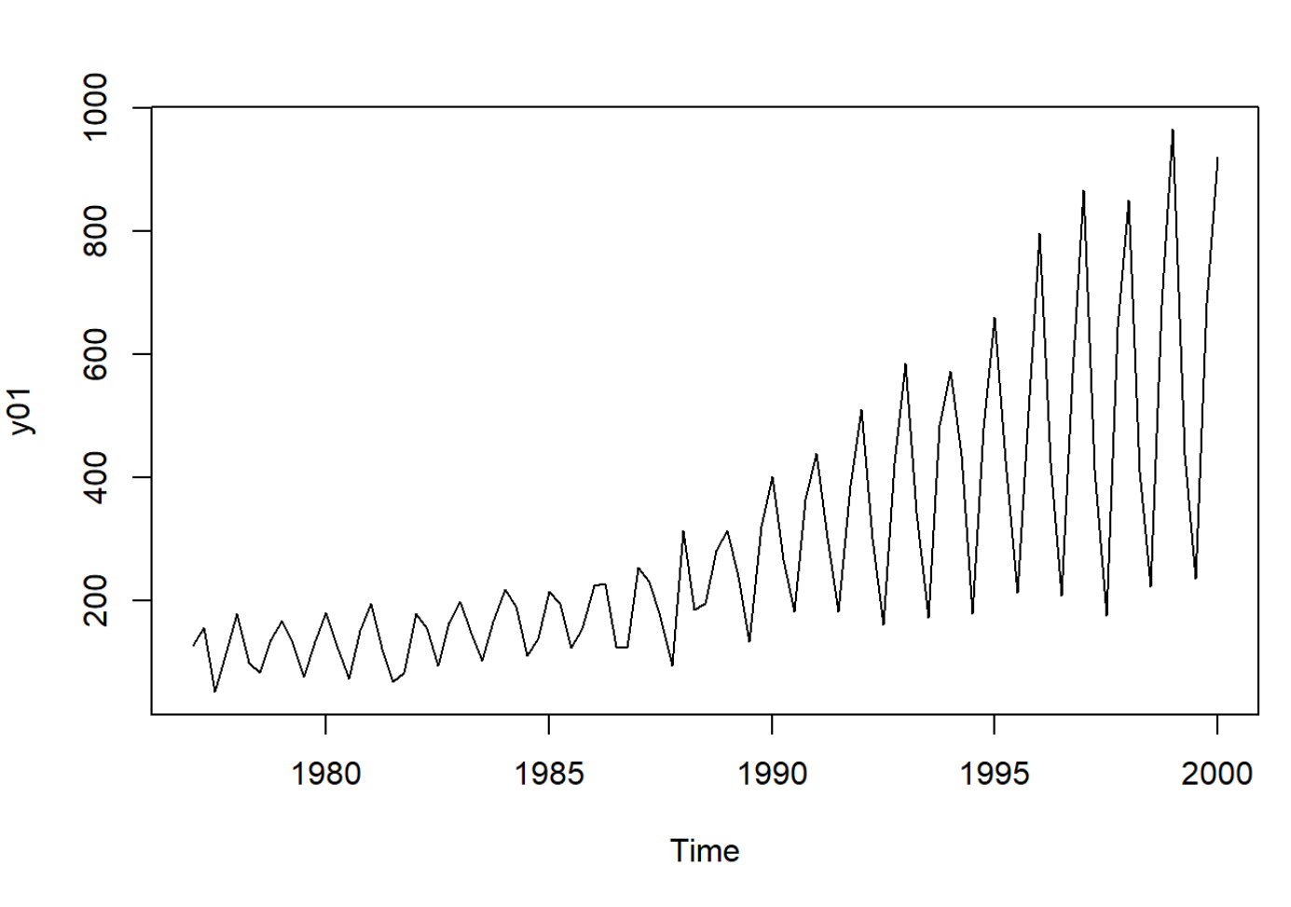
```
## Registered S3 methods overwritten by 'forecast':
##   method             from
##   fitted.Arima       TSA
##   fitted.fracdiff    fracdiff
##   plot.Arima         TSA
##   residuals.fracdiff fracdiff
```

```
library(astsa)
```

```
##
## Attaching package: 'astsa'
```

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

```
load(url("http://goo.gl/Zqn3Ea"))
plot(y01)
```



Part A

1)

```
mymodel = tslm(y01 ~ season)
s = summary(mymodel)
s
```

```
##
## Call:
## tslm(formula = y01 ~ season)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -303.71 -137.45  -22.14   84.48  534.86
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   430.65      38.01  11.330  < 2e-16 ***
## season2      -171.46      54.34  -3.156  0.00219 **
## season3      -285.21      54.34  -5.249 1.03e-06 ***
## season4      -113.93      54.34  -2.097  0.03886 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 186.2 on 89 degrees of freedom
## Multiple R-squared:  0.2436, Adjusted R-squared:  0.2181
## F-statistic: 9.552 on 3 and 89 DF,  p-value: 1.56e-05
```

A model has the form:

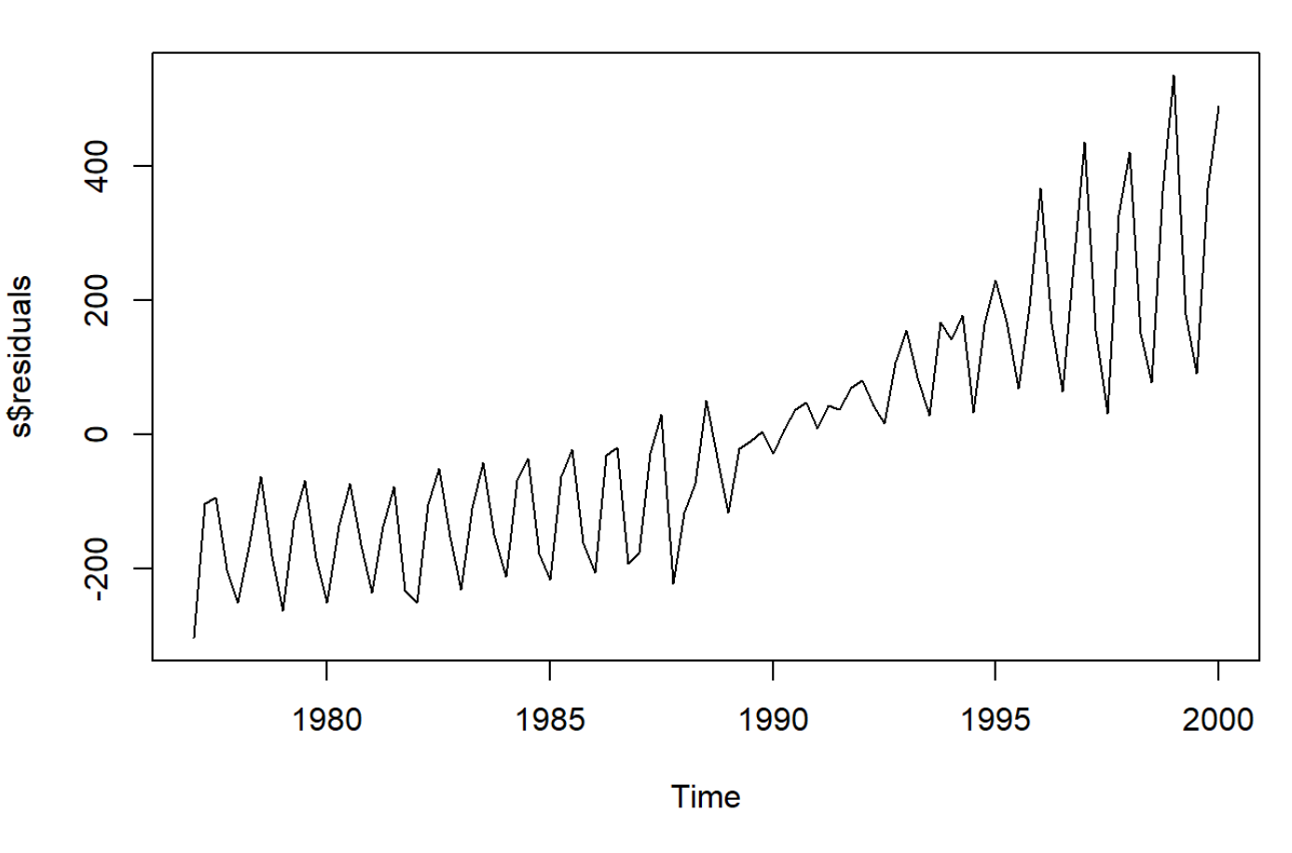
$$y(t) = 430.65 - 171.46 * season2 * x_{(2,t)} - 285.21 * season3 * x_{(3,t)} - 113.93 * season4 * x_{(4,t)}$$

Where season2,season3,season4 are dummy variables (indicators of event) of quarterly 2,3, and 4 respectively.

2)

Plot of residuals:

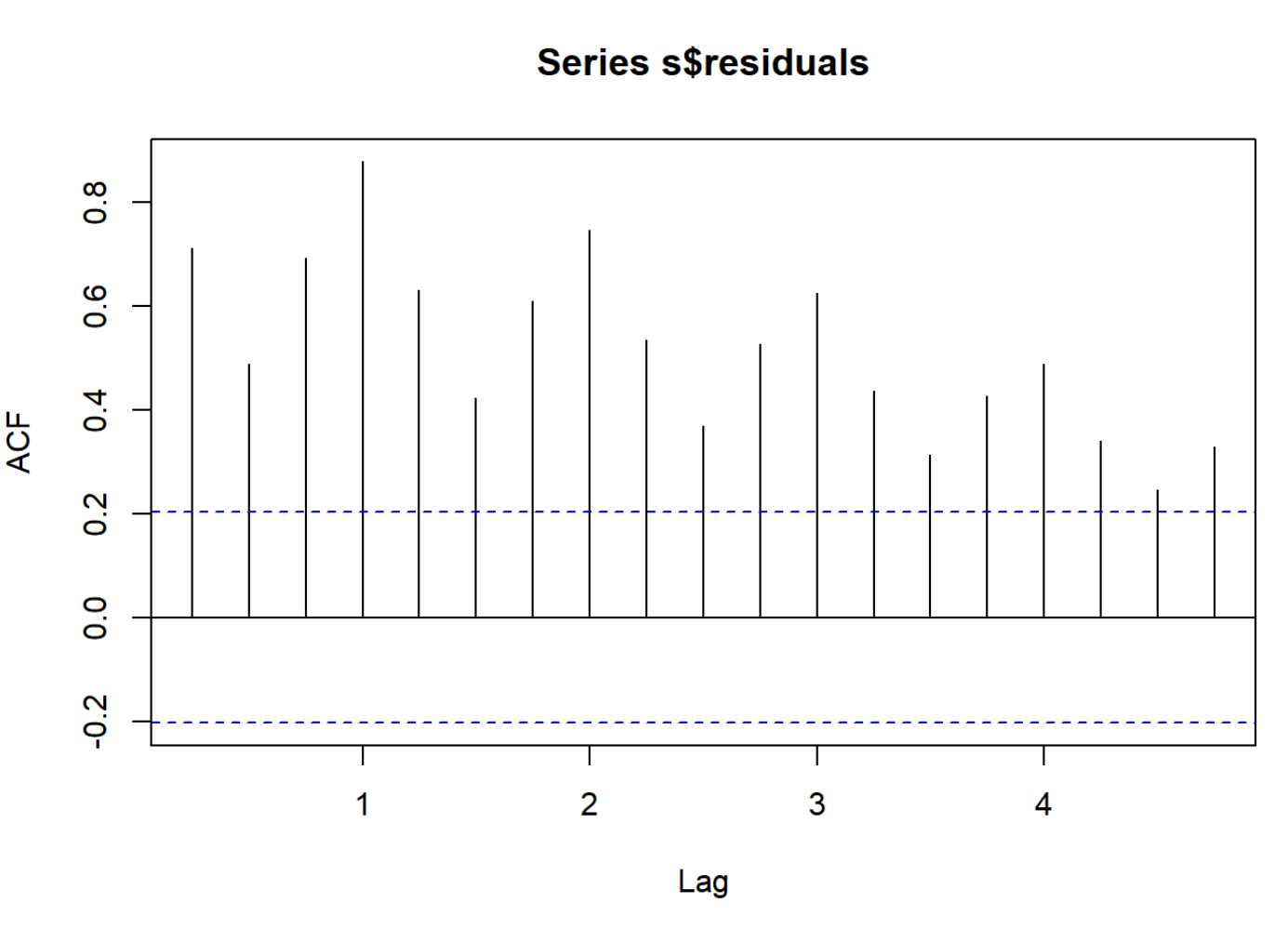
```
plot(s$residuals)
```



The residuals are non-uniform with respect to zero, therefore only seasonality is not enough to forecast the time series.

ACF:

```
acf(s$residuals)
```



The residuals are correlated.

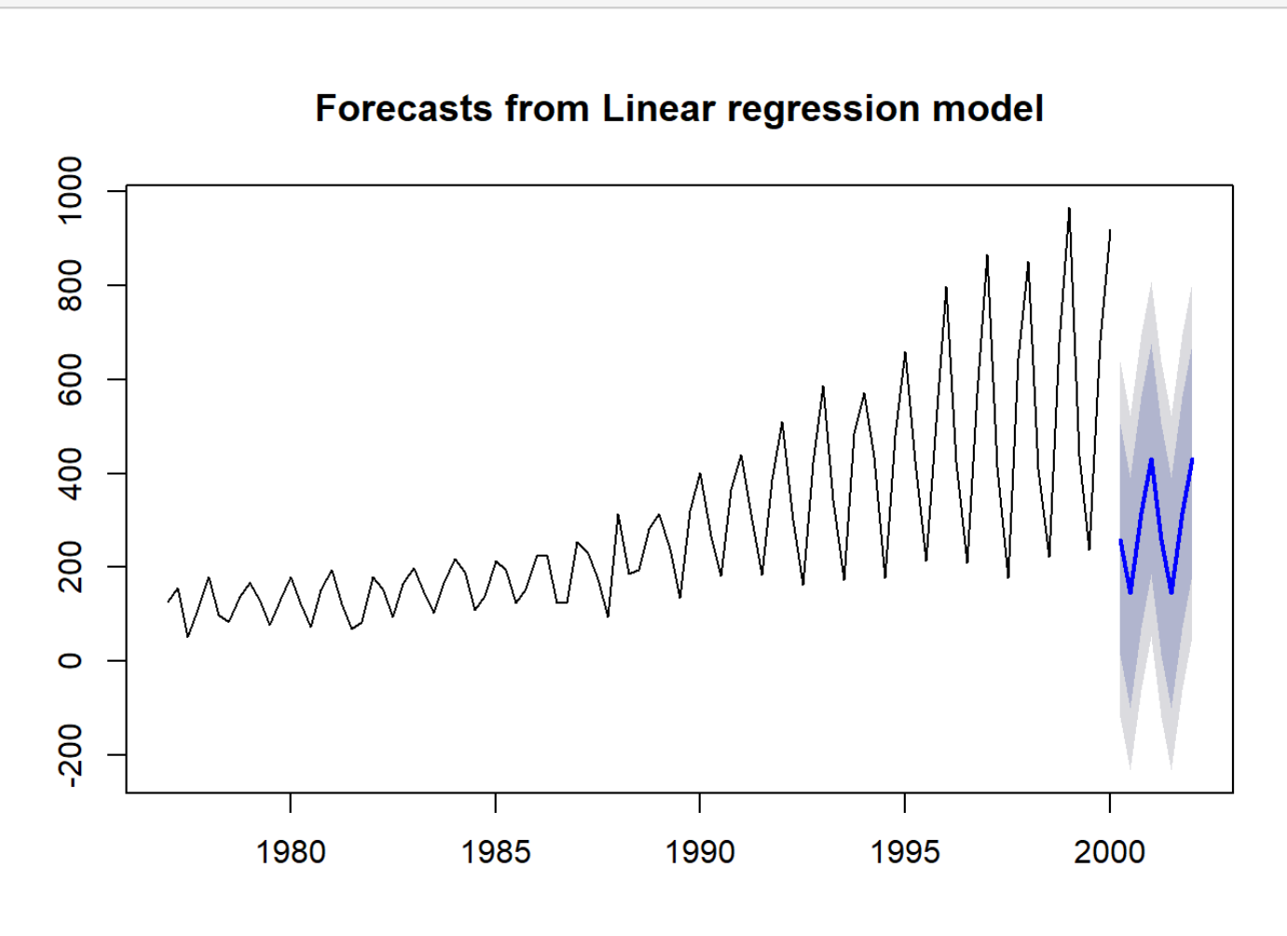
We can conclude that we got a rather bad model. This is also seen from the forecast graph below:

3)

```
forecast(mymodel, h = 8)
```

```
##         Point Forecast      Lo 80    Hi 80      Lo 95    Hi 95
## 2000 Q2       259.1908   13.59674 504.7849 -118.76327 637.1449
## 2000 Q3       145.4417 -100.15239 391.0358 -232.51240 523.3958
## 2000 Q4       316.7215   71.12745 562.3156  -61.23256 694.6756
## 2001 Q1       430.6478  185.26695 676.0286   53.02189 808.2736
## 2001 Q2       259.1908   13.59674 504.7849 -118.76327 637.1449
## 2001 Q3       145.4417 -100.15239 391.0358 -232.51240 523.3958
## 2001 Q4       316.7215   71.12745 562.3156  -61.23256 694.6756
## 2002 Q1       430.6478  185.26695 676.0286   53.02189 808.2736
```

```
plot(forecast(mymodel, h = 8))
```



Part B

1)

```
mymodel = arima(y01, order = c(1,1,1), seasonal = list(order = c(1,1,0), period = 4))
mymodel
```

```
##
## Call:
## arima(x = y01, order = c(1, 1, 1), seasonal = list(order = c(1, 1, 0), period = 4))
##
## Coefficients:
##           ar1      ma1     sar1
##       -0.1505  -0.9312  -0.0954
## s.e.   0.1112   0.0378   0.1121
##
## sigma^2 estimated as 1675:  log likelihood = -452.75,  aic = 911.49
```

A model has the form:

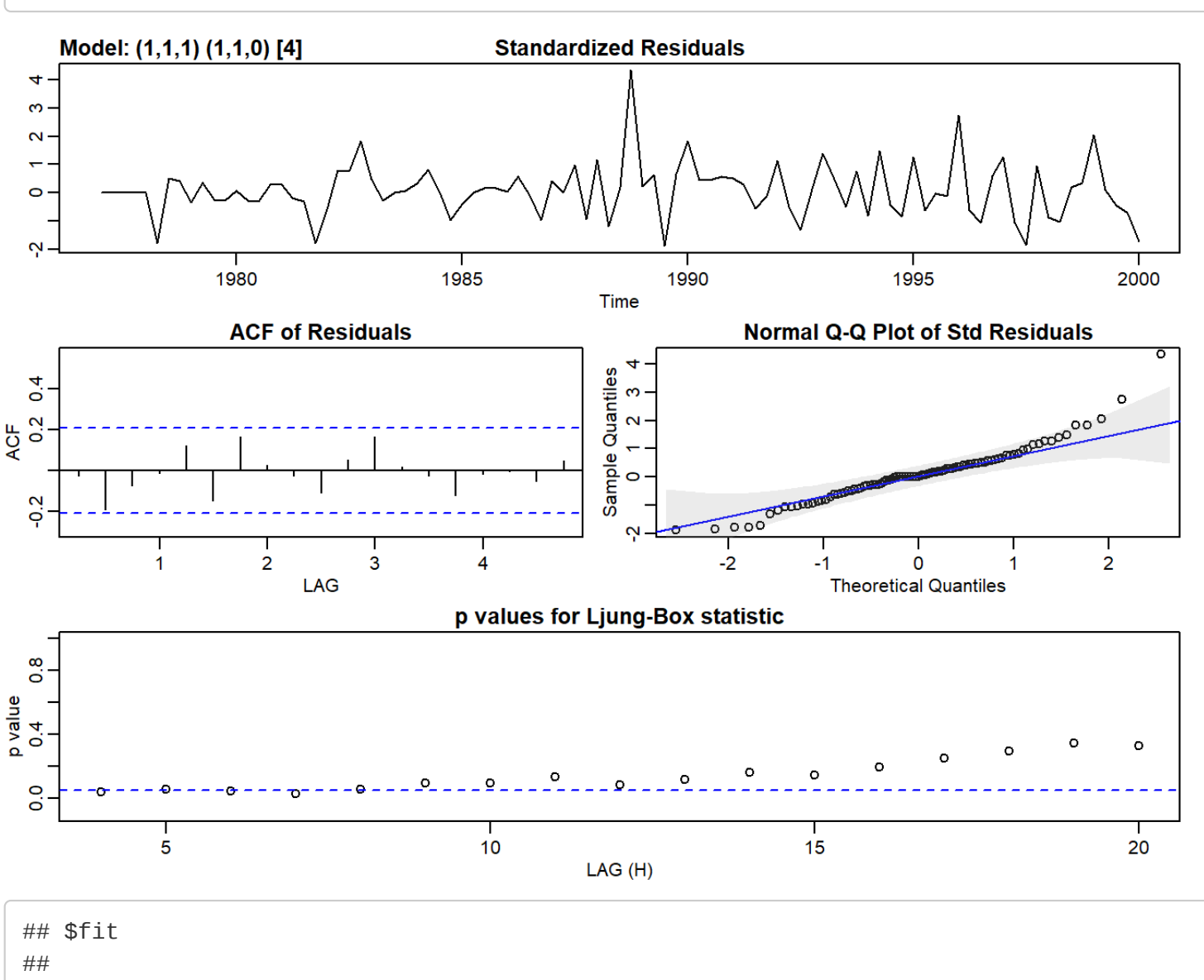
$$(1 - 0.1505L)(1 - L)(1 - L^4)(1 - 0.0954L^4)y(t) = \varepsilon_t - 0.9312\varepsilon_{(t-1)}$$

where L is Lag-operator.

2)

```
sarima(y01, 1,1,1,1,1,0,4)
```

```
## initial  value 4.101508
## iter   2 value 3.870196
## iter   3 value 3.777895
## iter   4 value 3.759068
## iter   5 value 3.746634
## iter   6 value 3.732811
## iter   7 value 3.723425
## iter   8 value 3.723318
## iter   9 value 3.723002
## iter  10 value 3.722990
## iter  11 value 3.722989
## iter  12 value 3.722989
## iter  12 value 3.722989
## iter  12 value 3.722989
## final  value 3.722989
## converged
## initial  value 3.727727
## iter   2 value 3.727684
## iter   3 value 3.725999
## iter   4 value 3.725918
## iter   5 value 3.725913
## iter   5 value 3.725913
## iter   5 value 3.725913
## final  value 3.725913
## converged
```



```
## $fit
##
## Call:
## stats::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 = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##           ar1      ma1     sar1
##       -0.1505  -0.9312  -0.0954
## s.e.   0.1112   0.0378   0.1121
##
## sigma^2 estimated as 1675:  log likelihood = -452.75,  aic = 913.49
##
## $degrees_of_freedom
## [1] 85
##
## $ttable
##      Estimate     SE  t.value p.value
## ar1   -0.1505 0.1112  -1.3530  0.1796
## ma1   -0.9312 0.0378 -24.6289  0.0000
## sar1  -0.0954 0.1121  -0.8512  0.3970
##
## $AIC
## [1] 10.03839
##
## $AICc
## [1] 10.04143
##
## $BIC
## [1] 10.14729
```

From the diagnostic graphs, it can be seen that the residues are evenly distributed relative to zero, do not correlate with each other (on the ACF plot, the blue lines do not intersect the black dashed lines), they are almost normally distributed (QC graph).

But on the plot of p-values of the test at points 4, 6, 7, we see the dependence in this time series.

3)

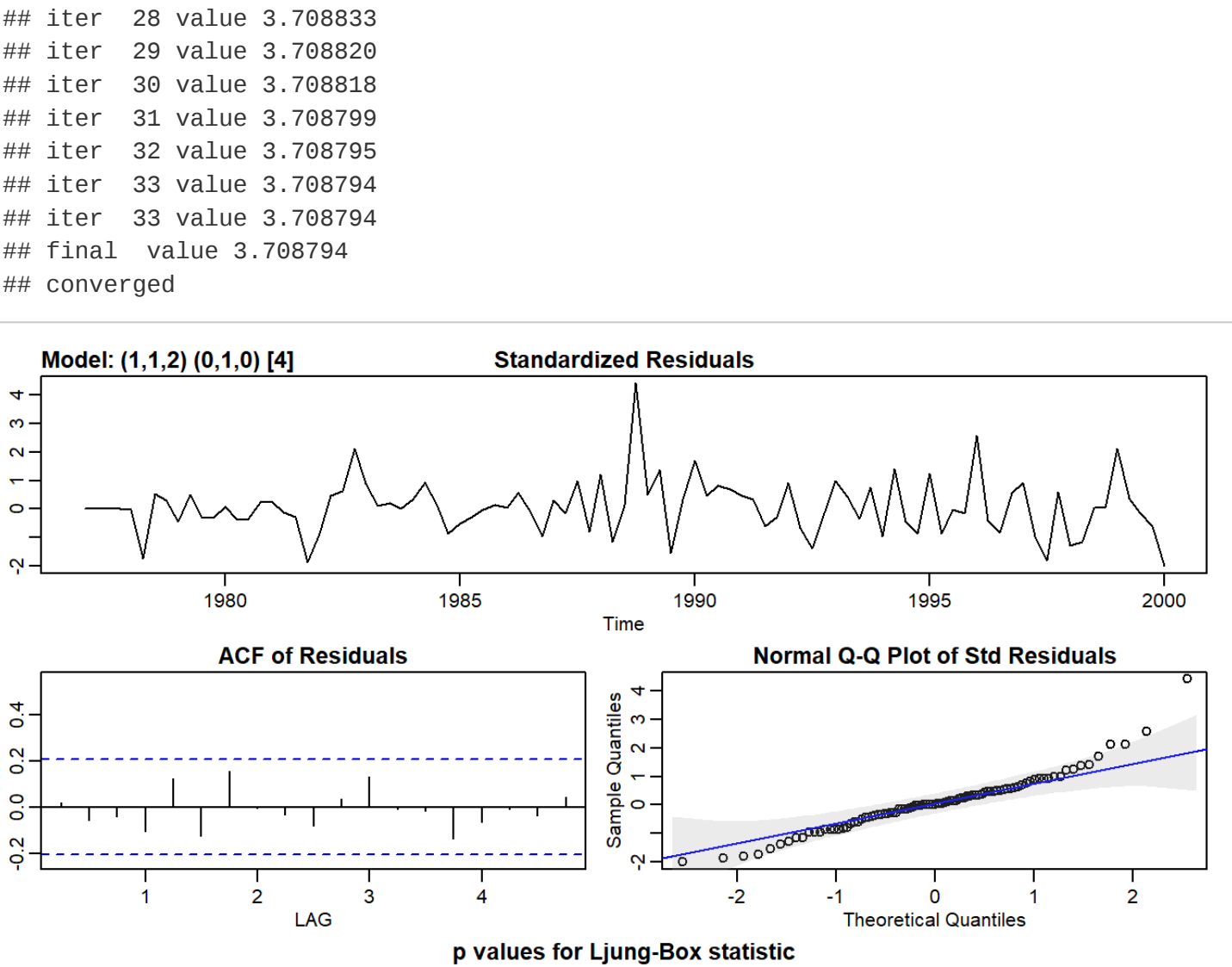
The model of part B is better than the model of part A, but it can be improved by choosing the right parameters. One solution to this problem is to use the auto.arima() function

```
auto.arima(y01,max.p = 10,
           max.q = 10,
           max.P = 10,
           max.Q = 10,
           max.order = 10,
           max.d = 10,
           max.D = 1,
           start.p = 0,
           start.q = 0,
           start.P = 0,
           start.Q = 0)
```

```
## Series: y01
## ARIMA(1,1,2)(0,1,0)[4]
##
## Coefficients:
##          ar1      ma1     ma2
##       0.4538  -1.6307  0.6648
## s.e.  0.2911   0.2651  0.2548
##
## sigma^2 estimated as 1671:  log likelihood=-451.24
## AIC=910.48   AICc=910.96   BIC=920.39
```

```
mymodel1 = sarima(y01, 1,1,2,0,1,0,4)
```

```
## initial  value 4.095531
## iter   2 value 3.851306
## iter   3 value 3.800430
## iter   4 value 3.775081
## iter   5 value 3.766119
## iter   6 value 3.754070
## iter   7 value 3.743964
## iter   8 value 3.742564
## iter   9 value 3.742166
## iter  10 value 3.740582
## iter  11 value 3.734386
## iter  12 value 3.733451
## iter  13 value 3.733093
## iter  14 value 3.733024
## iter  15 value 3.733012
## iter  15 value 3.733012
## final  value 3.733012
## converged
## initial  value 3.743154
## iter   2 value 3.741420
## iter   3 value 3.740849
## iter   4 value 3.740752
## iter   5 value 3.740740
## iter   6 value 3.740708
## iter   7 value 3.740519
## iter   8 value 3.740437
## iter   9 value 3.739239
## iter  10 value 3.738196
## iter  11 value 3.738071
## iter  12 value 3.720726
## iter  13 value 3.720151
## iter  14 value 3.718658
## iter  15 value 3.718014
## iter  16 value 3.716807
## iter  17 value 3.714008
## iter  18 value 3.712567
## iter  19 value 3.710484
## iter  20 value 3.710120
## iter  21 value 3.709396
## iter  22 value 3.709070
## iter  23 value 3.708925
## iter  24 value 3.708913
## iter  25 value 3.708887
## iter  26 value 3.708850
## iter  27 value 3.708850
## iter  28 value 3.708833
## iter  29 value 3.708820
## iter  30 value 3.708818
## iter  31 value 3.708799
## iter  32 value 3.708795
## iter  33 value 3.708794
## iter  33 value 3.708794
## final  value 3.708794
## converged
```



```
mymodel1
```

```
## $fit
##
## Call:
## stats::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 = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ma1     ma2
##       0.4538  -1.6307  0.6648
## s.e.  0.2911   0.2651  0.2548
##
## sigma^2 estimated as 1614:  log likelihood = -451.24,  aic = 910.48
##
## $degrees_of_freedom
## [1] 85
##
## $ttable
##     Estimate     SE t.value p.value
## ar1   0.4538 0.2911  1.5591  0.1227
## ma1  -1.6307 0.2651 -6.1525  0.0000
## ma2   0.6648 0.2548  2.6097  0.0107
##
## $AIC
## [1] 10.00529
##
## $AICc
## [1] 10.00832
##
## $BIC
## [1] 10.11418
```

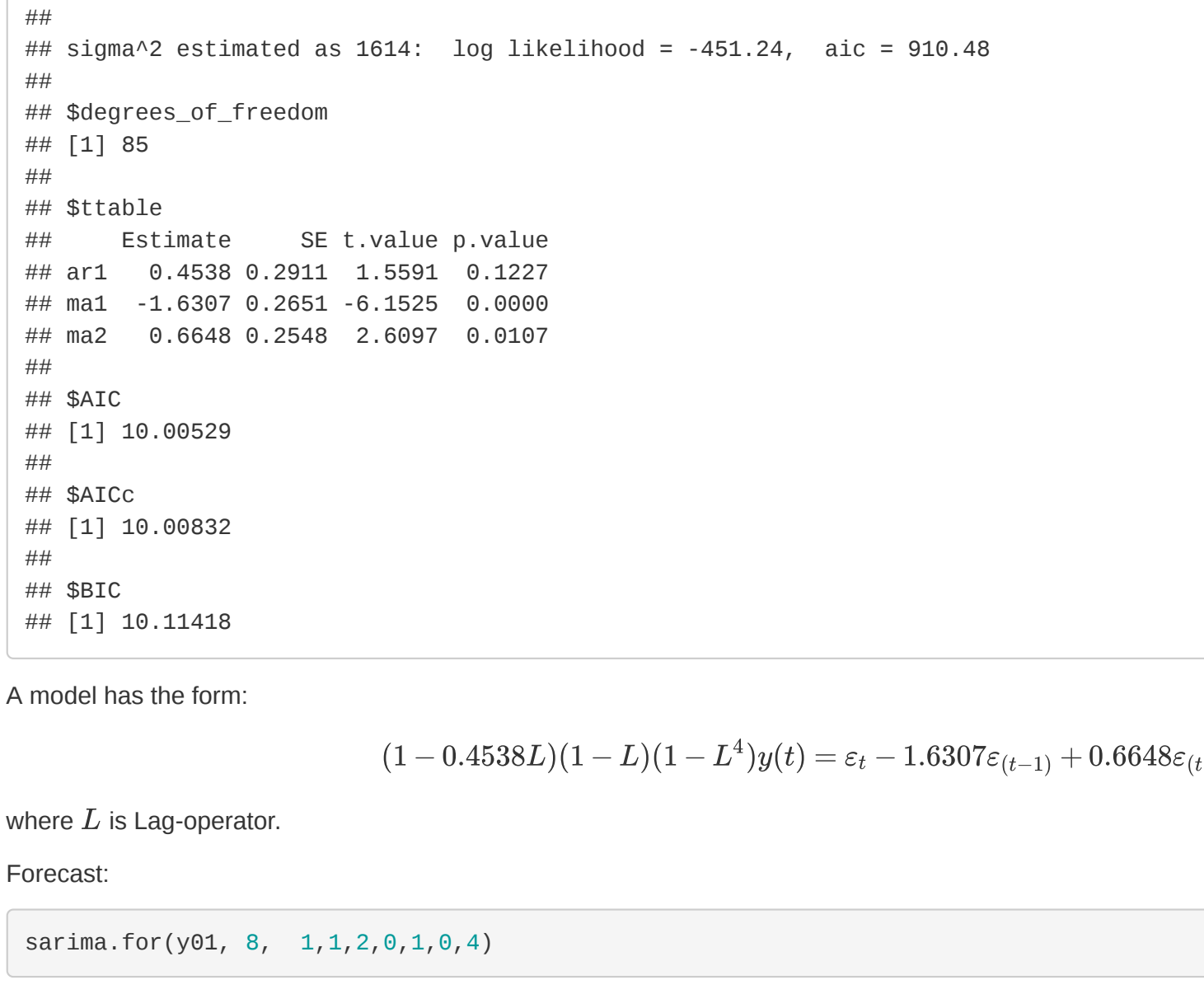
A model has the form:

$$(1 - 0.4538L)(1 - L)(1 - L^4)y(t) = \varepsilon_t - 1.6307\varepsilon_{(t-1)} + 0.6648\varepsilon_{(t-2)}$$

where L is Lag-operator.

Forecast:

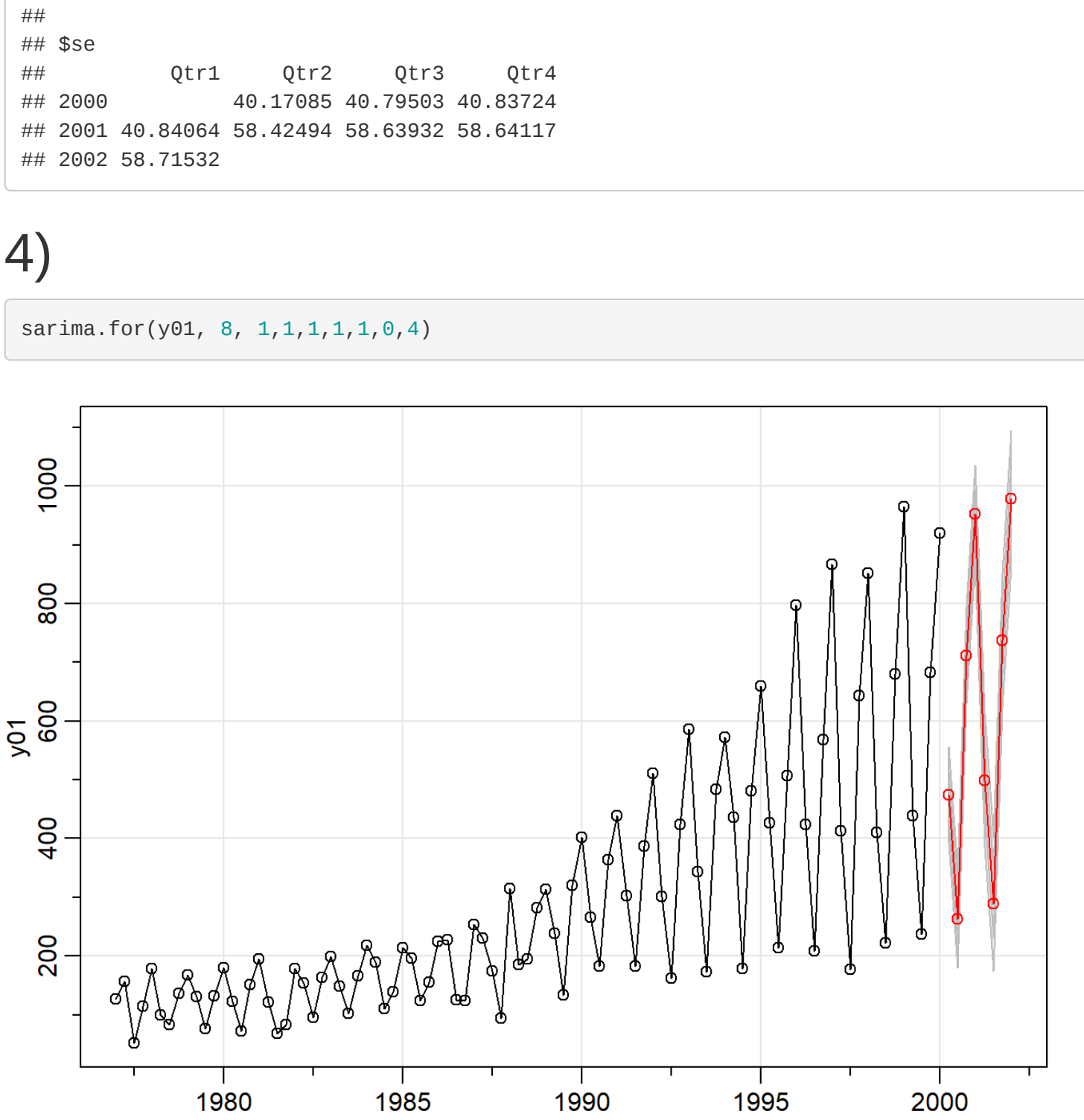
```
sarima.for(y01, 8,  1,1,2,0,1,0,4)
```



```
## $pred
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2000          485.1280 271.8816 713.4387
## 2001 948.1090 511.9485 298.2147 739.5507
## 2002 974.1206
##
## $se
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2000          40.17085 40.79503 40.83724
## 2001 40.84064 58.42494 58.63932 58.64117
## 2002 58.71532
```

4)

```
sarima.for(y01, 8, 1,1,1,1,1,0,4)
```



```
## $pred
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2000          473.6559 262.3243 711.7509
## 2001 953.2184 498.9857 288.5629 737.7371
## 2002 978.7816
##
## $se
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2000          40.92877 41.06489 41.19891
## 2001 41.26407 57.11634 57.11924 57.37894
## 2002 57.55873
```

Part C

Of course, the model of part B is much better than the model of part A. This can be seen both in the forecast graph and in the diagnostic graphs of residuals.