

2º Trabalho - Série Temporais

Parte III: 10, 18, 20, 31 e 36. Parte IV: 6

Milton Abraao Ferreira / GRR20137581

2019-12-02

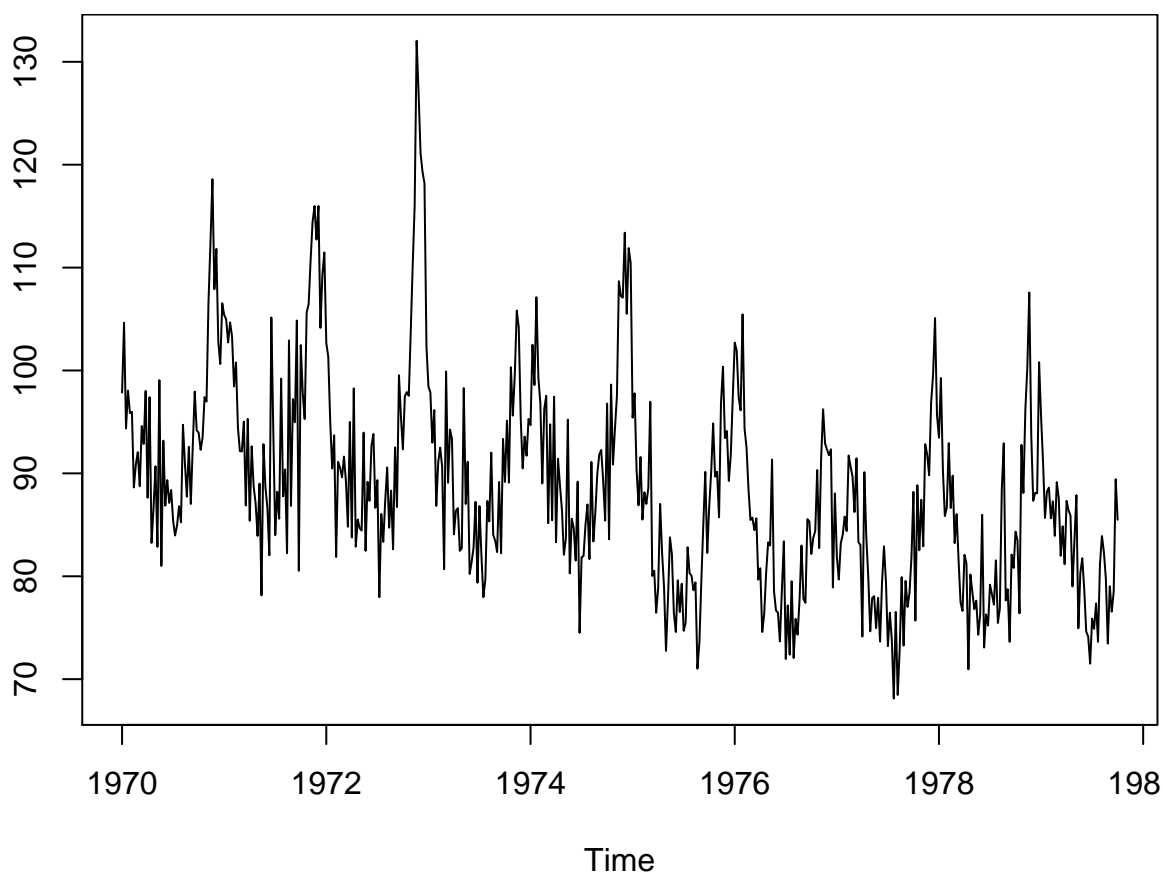
PARTE III - Modelos ARIMA

Exercício 10

- a) Com os dados *cmort* foi ajustado um AR(2) usando regressão linear. Os dados foram coletados semanalmente ao longo dos anos, por isso a frequência utilizada para transformar os dados em uma série foi 52.

```
library(astsa)

cmort.ts = ts(cmort, freq = 52, start = c(1970,1))
par(mar=c(4,4,0,0.5));plot.ts(cmort.ts)
```



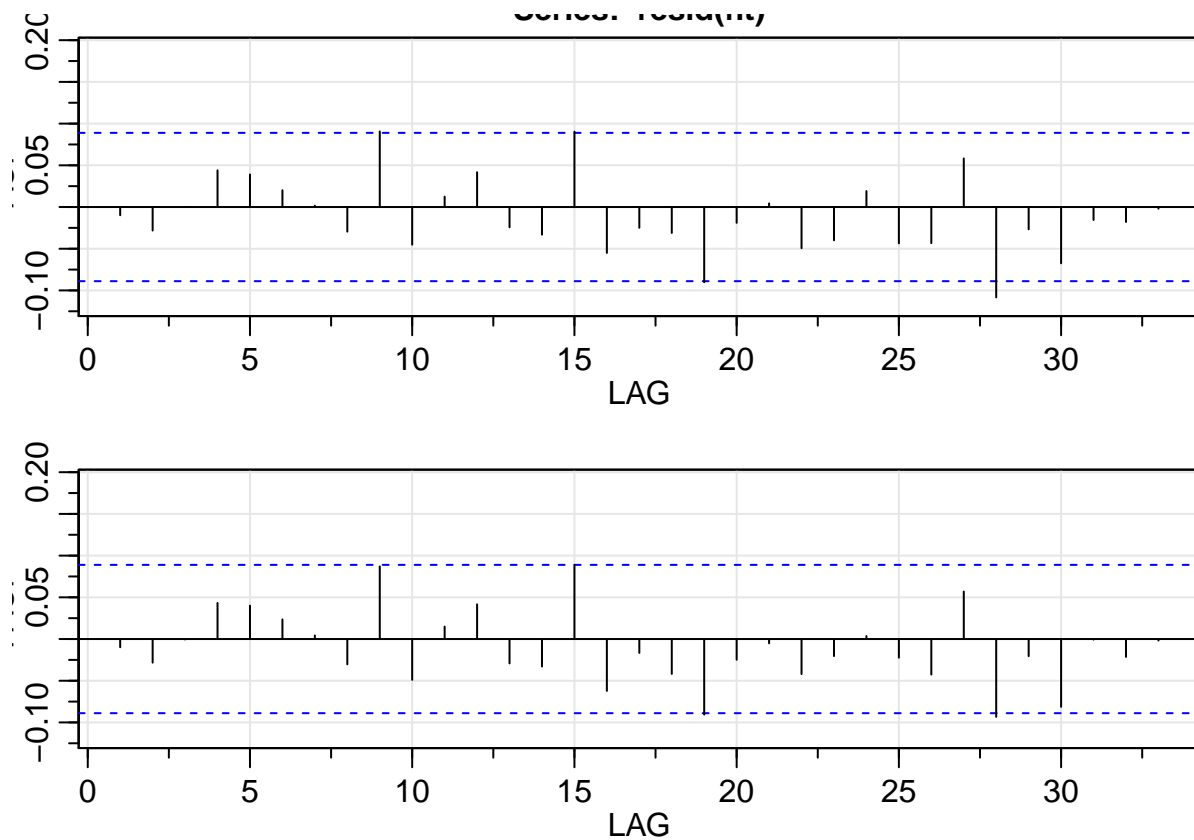
```
x.lag1 <- lag(cmort.ts, 1)
x.lag2 <- lag(cmort.ts, 2)

dbajuste = na.omit(cbind(cmort.ts,x.lag2, x.lag1))
```

```
fit <- lm(cmort.ts ~ ., data = dbajuste)
```

Pelo gráfico acima é possível observar a correlação dos erros, também na função de autocorrelação parcial (PACF) a correlação autoregressiva.

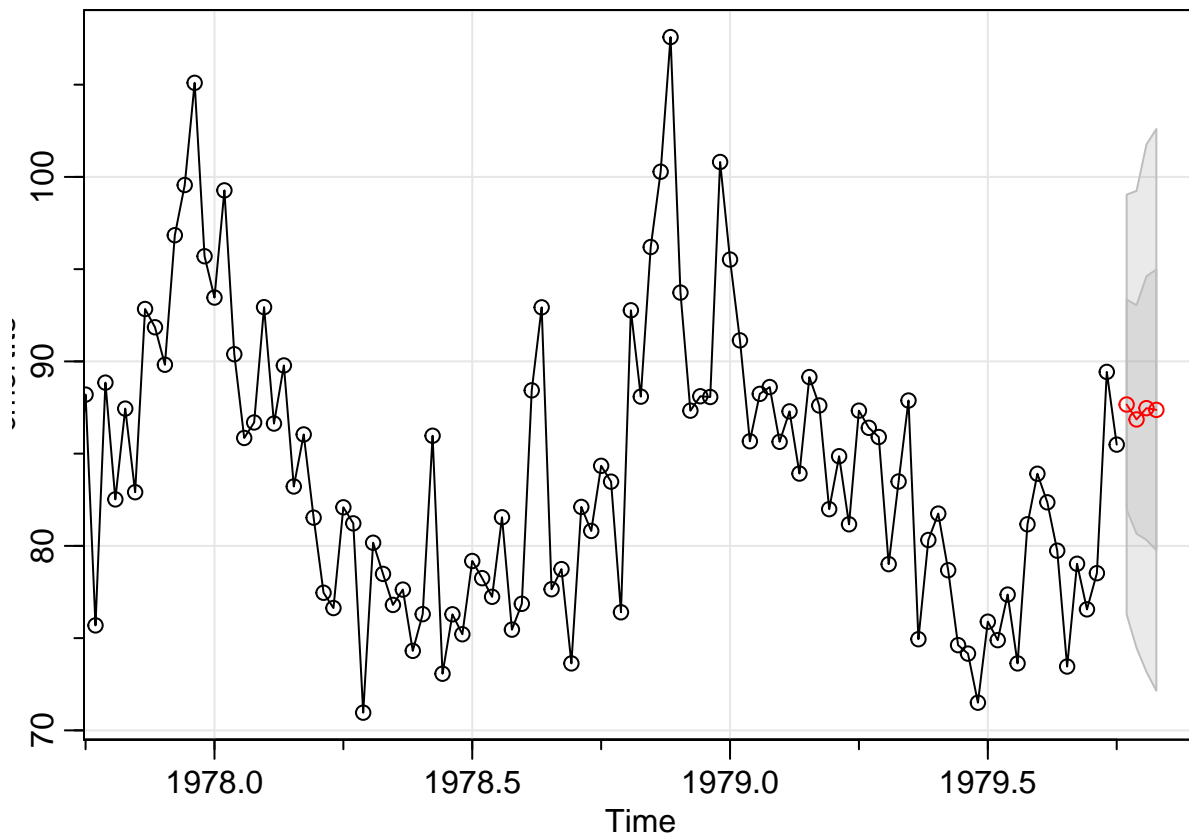
```
par(mar=c(4,4,0,0));acf2(resid(fit))
```



Como os resíduos estão dentro do intervalo de confiança, podemos considerar que o modelo está bem ajustado.

- b) Uma maneira alternativa de ajustar o modelo de regressão como feito no exercício acima é utilizando a função *sarima()*. Como neste caso estamos lidando com predição, o código utilizado é o *sarima.pred()*, assim como foi feito abaixo. Um ponto de atenção é que modelos lineares são mais sensíveis a grande números quando comparado ao ajuste com *sarima()*.

```
par(mar=c(4,4,0,0));sarima.for(cmort.ts, 2,0,0, n.ahead = 4)
```



O gráfico apresentou somente o final da série, uma vez que esse período possui os dados mais relevantes para o ajuste.

Exercício 18

Ajustando modelos para da questão (mesma base utilizada no exercício 10):

```
# Ajustando o AR(2)
AR2 <- arima(cmort, order=c(2,0,0))

# Usando regressão linear
mod1 <- ar.ols(cmort, order=2, demean=FALSE, intercept=TRUE)

# Usando o Yule-Walker
mod2 <- ar.yw(cmort, order=2)
```

- a) As estimativas ficaram muito próximas. A diferença é que a primeira possui estimativa para o intercepto, enquanto a segunda não.

```
# Usando regressão linear:
mod1$ar

## , , 1
##
##      [,1]
## [1,] 0.4285906
## [2,] 0.4417874
```

```
# Usando Yule-Walker:
mod2$ar
```

```
## [1] 0.4339481 0.4375768
```

- b) As estimativas dos erros padrões foram baixas nos dois modelos. Além de baixas, elas também foram bem próximas, reafirmando assim a “Proposição III.10” que diz respeito aos estimadores do modelo ARIMA para amostras grandes. Em outras palavras, muito provavelmente a distribuição dos estimadores para o modelo ARIMA está próxima da distribuição normal assintótica.

```
# Usando regressão linear:
mod1$asy.se.coef
```

```
## $x.mean
## [1] 2.393673
##
## $ar
## [1] 0.03979433 0.03976163
```

```
# Usando Yule-Walker:
mod2$ar
```

```
## [1] 0.4339481 0.4375768
```

```
sqrt(diag(mod2$asy.var.coef))
```

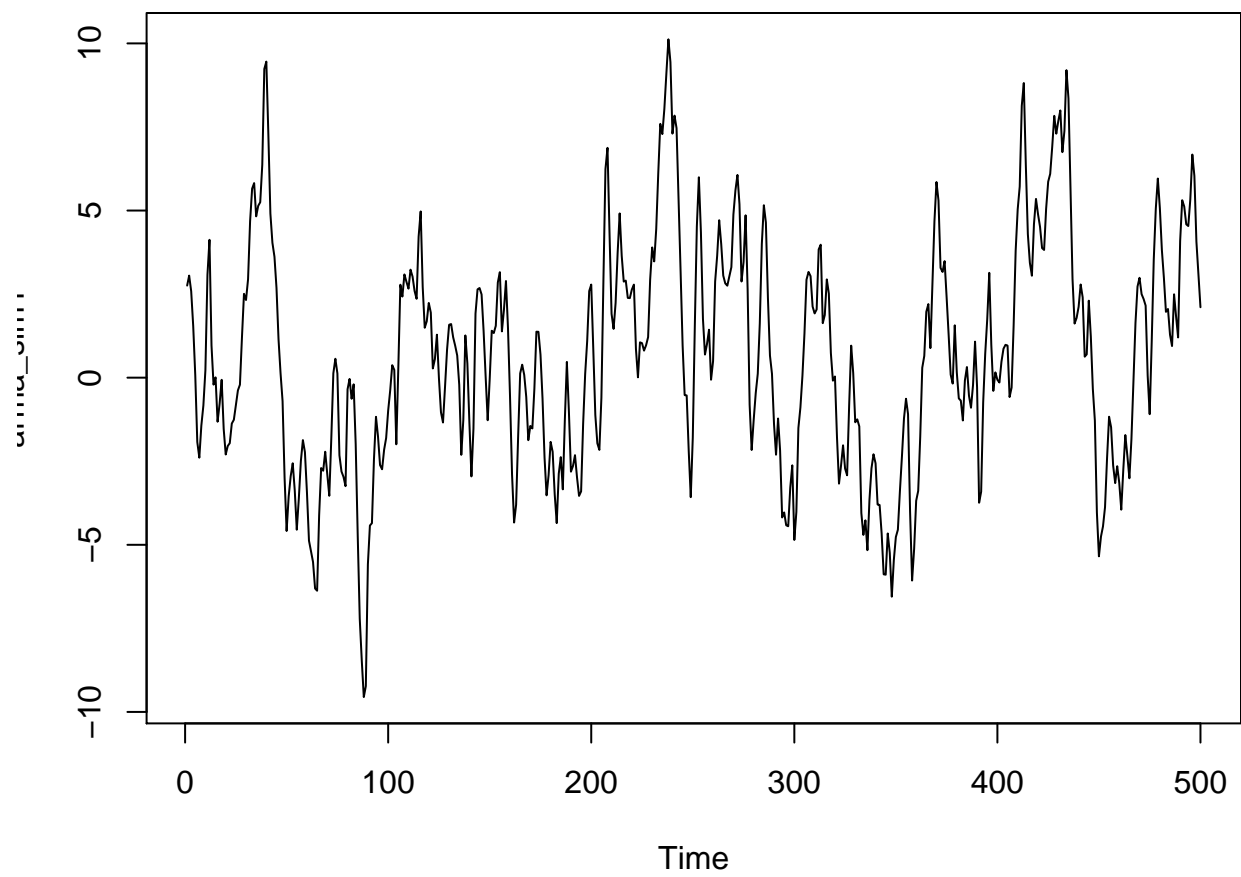
```
## [1] 0.04001303 0.04001303
```

Exercício 20

Os resultados diferem porque o modelo depende apenas dos dados amostrais realizados, as componentes autoregressivas e de medias móveis dependem apenas das ultimas realizações em cada modelo simulado, o que por si só define naturalmente em cada série.

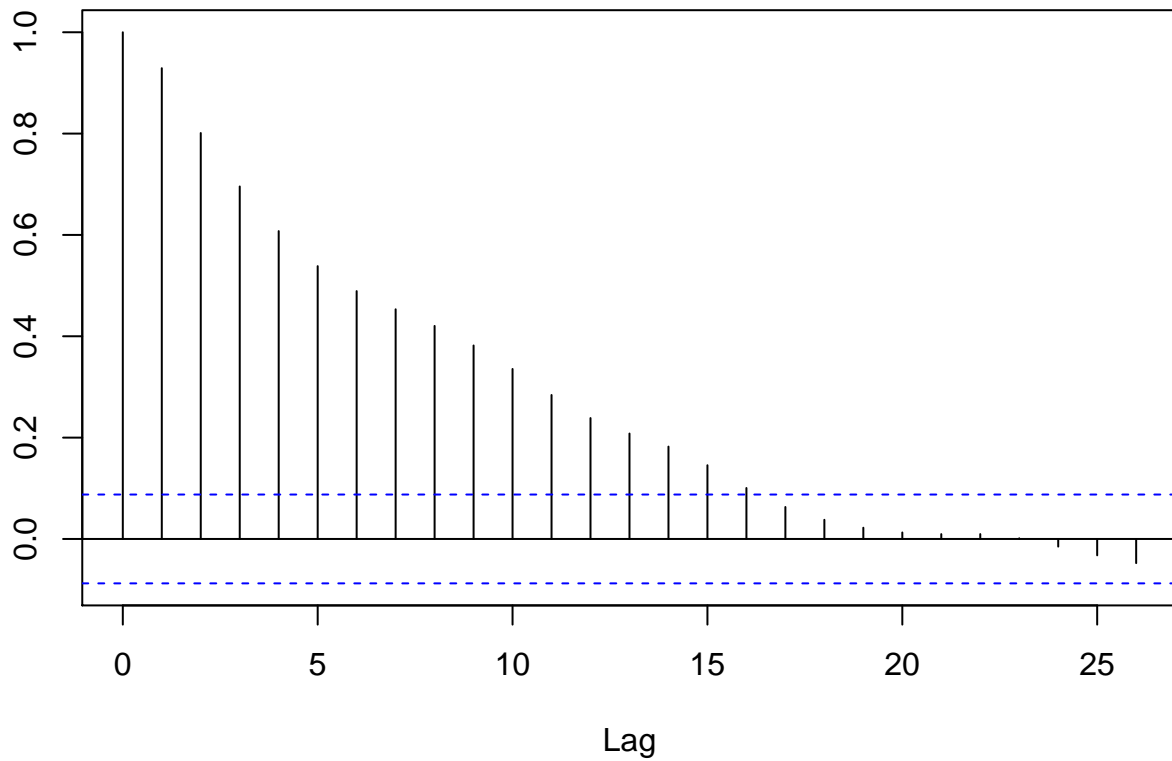
```
set.seed(123)
arma_sim1 <- arima.sim(model=list(ar=c(.9),ma=c(.9)),n=500)

par(mar=c(4,4,0,0));ts.plot(arma_sim1)
```



```
par(mar=c(4,4,3,0));arma_acf1 <- acf(arma_sim1,type="correlation",plot=T)
```

series 'arma_sim1'

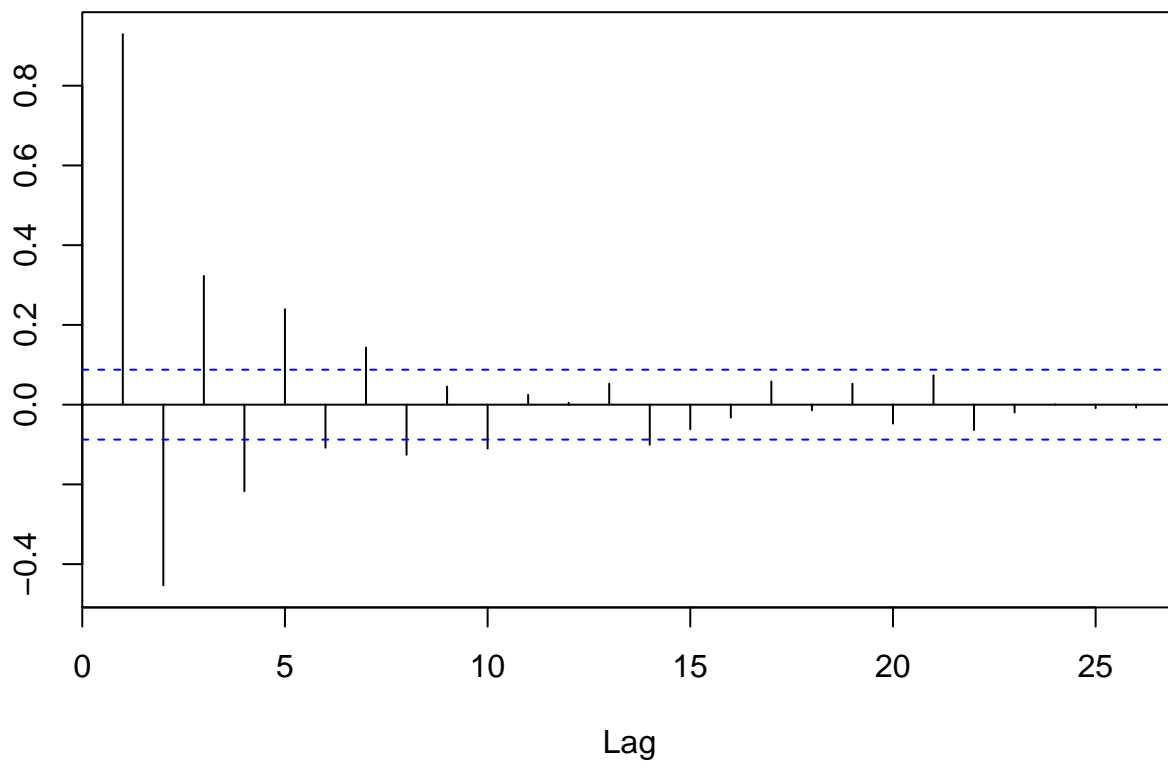


```
arma_acf1
```

```
##
## Autocorrelations of series 'arma_sim1', by lag
##
##      0      1      2      3      4      5      6      7      8      9
## 1.000 0.929 0.801 0.696 0.608 0.538 0.489 0.453 0.420 0.382
##    10    11    12    13    14    15    16    17    18    19
## 0.335 0.284 0.239 0.208 0.182 0.145 0.101 0.063 0.038 0.022
##    20    21    22    23    24    25    26
## 0.013 0.010 0.010 0.001 -0.015 -0.032 -0.048
```

```
par(mar=c(4,4,3,0));arma_pacf1 <-acf(arma_sim1,type="partial",plot=T)
```

series arma_sim1



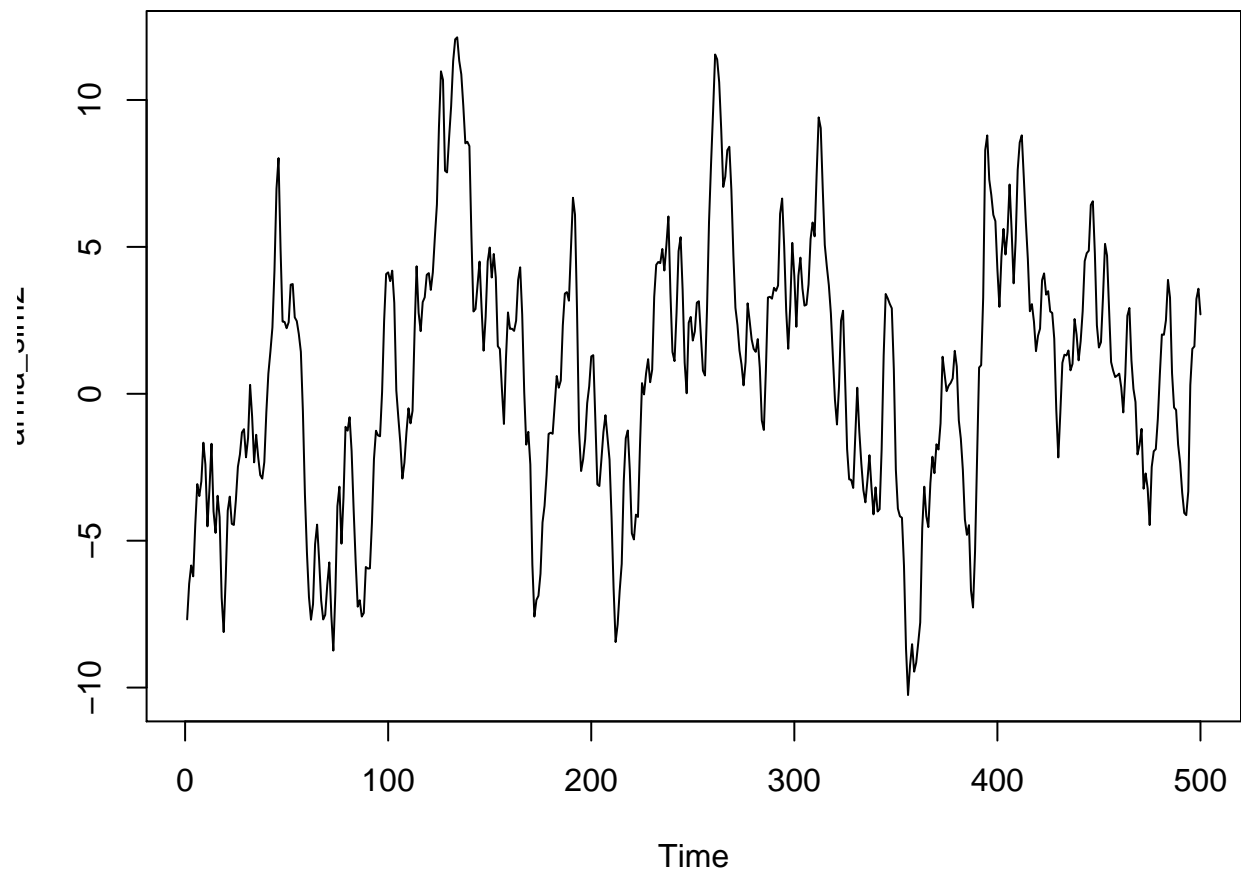
arma_pacf1

```
##
## Partial autocorrelations of series 'arma_sim1', by lag
##
##      1      2      3      4      5      6      7      8      9     10
## 0.929 -0.453 0.323 -0.217 0.239 -0.108 0.143 -0.126 0.045 -0.110
##      11     12     13     14     15     16     17     18     19     20
## 0.025 0.005 0.053 -0.101 -0.062 -0.032 0.058 -0.014 0.052 -0.048
##      21     22     23     24     25     26
## 0.074 -0.064 -0.020 0.001 -0.009 -0.007
```

```
ajust1 <- arima(arma_sim1, order = c(1,0,1), seasonal = c(0,0,0))
```

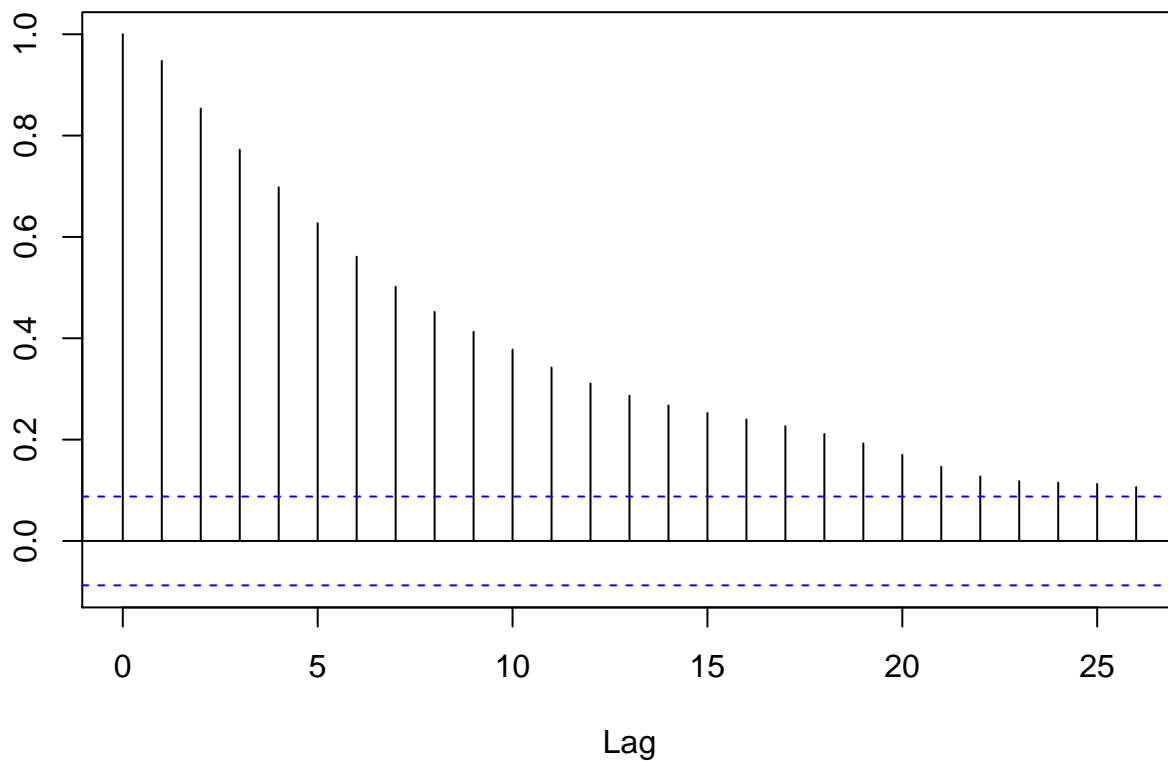
```
arma_sim2 <- arima.sim(model=list(ar=c(.9),ma=c(.9)),n=500)
```

```
par(mar=c(4,4,0,0));ts.plot(arma_sim2)
```



```
par(mar=c(4,4,3,0));arma_acf2 <-acf(arma_sim2,type="correlation",plot=T)
```


series arma_sim2

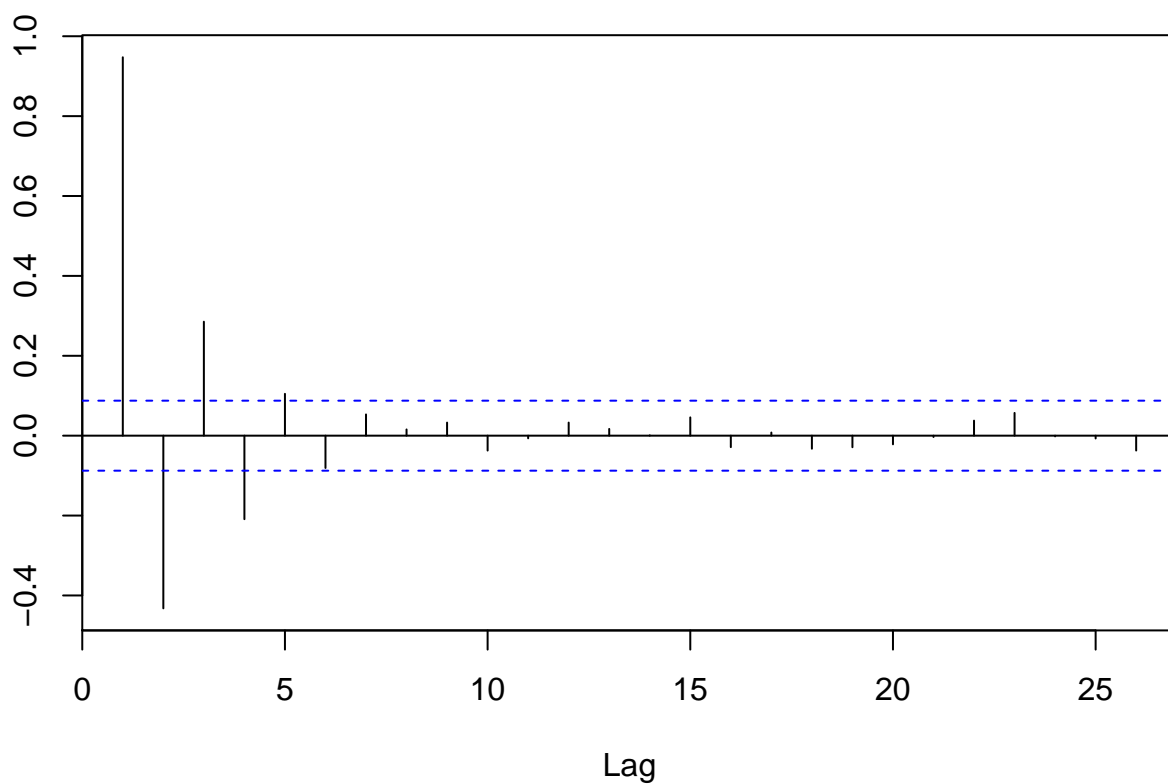


```
arma_acf2
```

```
##
## Autocorrelations of series 'arma_sim2', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10     11
## 1.000 0.947 0.853 0.772 0.698 0.627 0.561 0.502 0.452 0.413 0.378 0.342
##     12     13     14     15     16     17     18     19     20     21     22     23
## 0.311 0.287 0.267 0.252 0.240 0.226 0.211 0.192 0.170 0.146 0.127 0.118
##     24     25     26
## 0.115 0.113 0.106
```

```
par(mar=c(4,4,3,0));arma_pacf2 <-acf(arma_sim2,type="partial",plot=T)
```

series arma_sim2



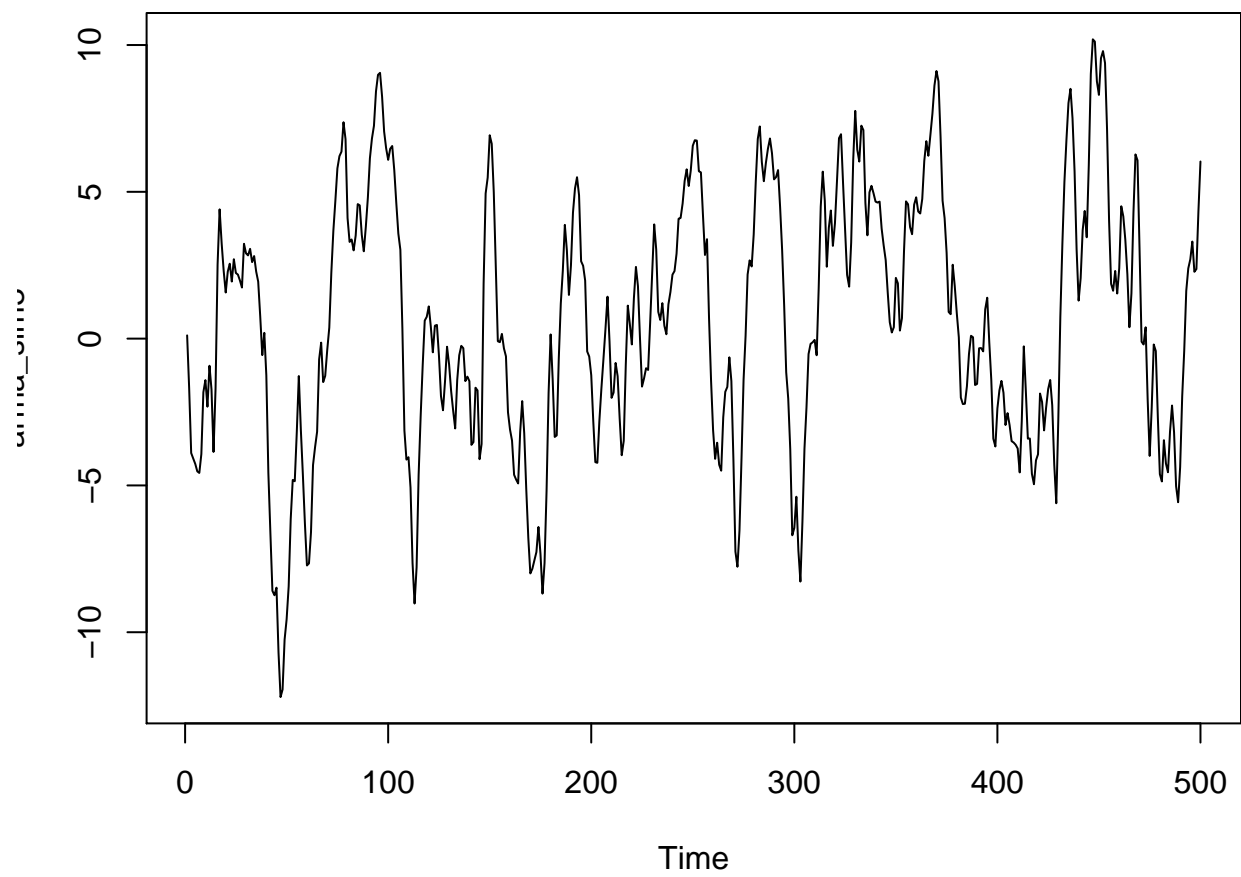
arma_pacf2

```
##
## Partial autocorrelations of series 'arma_sim2', by lag
##
##      1      2      3      4      5      6      7      8      9     10
## 0.947 -0.432 0.285 -0.209 0.105 -0.081 0.053 0.016 0.033 -0.037
##      11     12     13     14     15     16     17     18     19     20
## -0.007 0.033 0.017 0.001 0.046 -0.029 0.008 -0.033 -0.029 -0.022
##      21     22     23     24     25     26
## -0.003 0.038 0.057 -0.002 -0.007 -0.037
```

```
ajust2 <- arima(arma_sim2, order = c(1,0,1), seasonal = c(0,0,0))
```

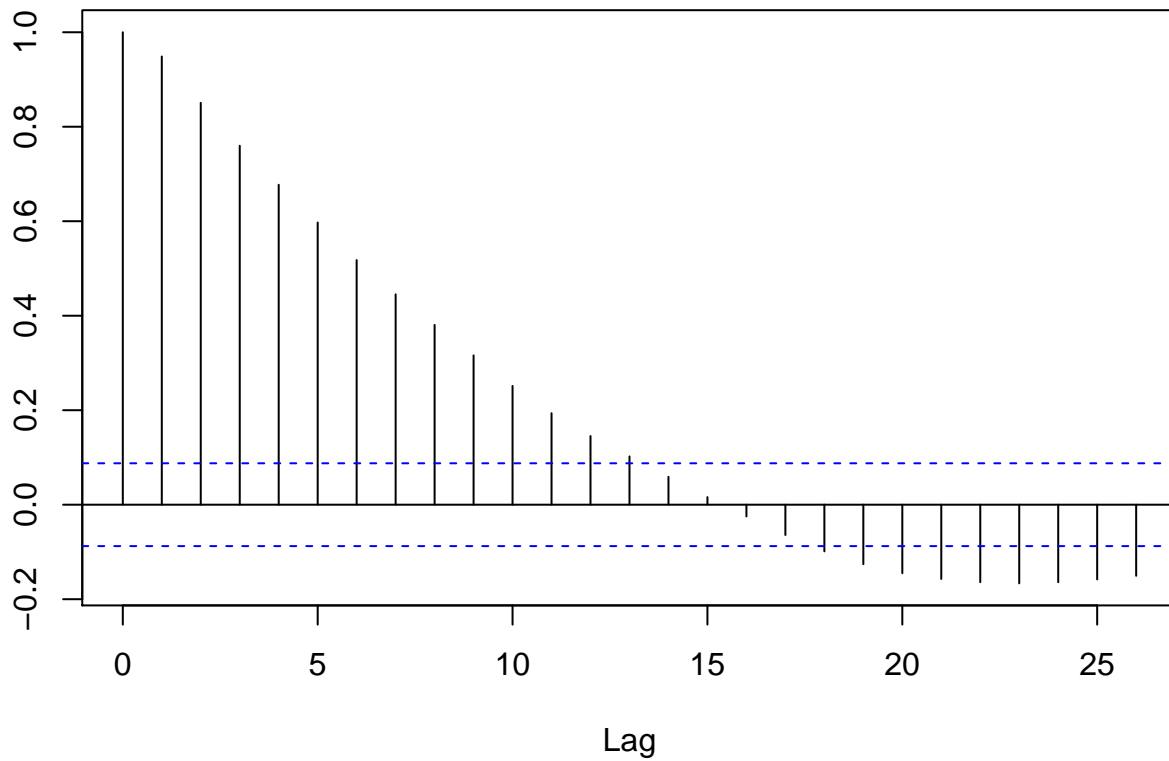
```
arma_sim3 <- arima.sim(model=list(ar=c(.9),ma=c(.9)),n=500)
```

```
par(mar=c(4,4,0,0));ts.plot(arma_sim3)
```



```
par(mar=c(4,4,3,0));arma_acf3 <- acf(arma_sim3,type="correlation",plot=T)
```

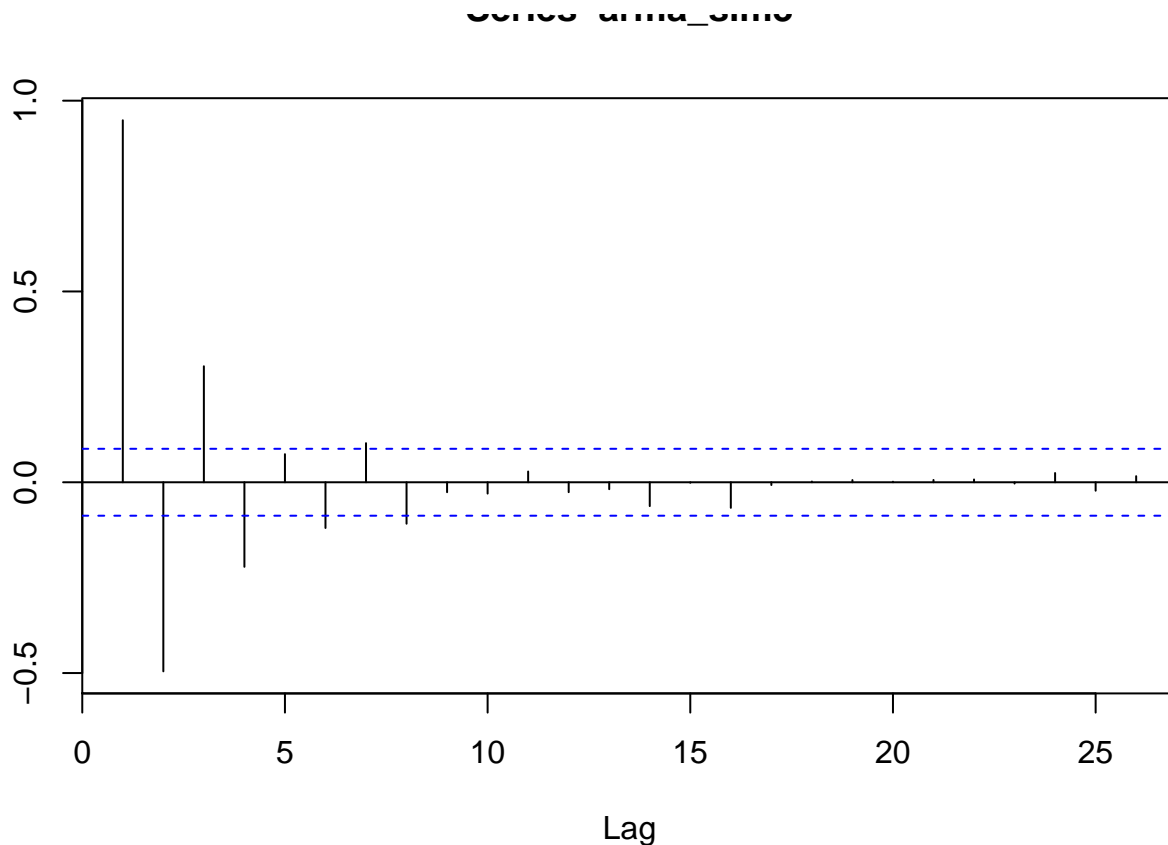
series arma_sim3



```
arma_acf3
```

```
##
## Autocorrelations of series 'arma_sim3', by lag
##
##      0      1      2      3      4      5      6      7      8      9
## 1.000 0.949 0.851 0.760 0.677 0.597 0.518 0.445 0.381 0.316
##    10    11    12    13    14    15    16    17    18    19
## 0.251 0.194 0.145 0.102 0.059 0.016 -0.025 -0.064 -0.099 -0.126
##    20    21    22    23    24    25    26
## -0.145 -0.157 -0.164 -0.167 -0.164 -0.158 -0.150
```

```
par(mar=c(4,4,3,0));arma_pacf3<-acf(arma_sim3,type="partial",plot=T)
```



```
arma_pacf3
```

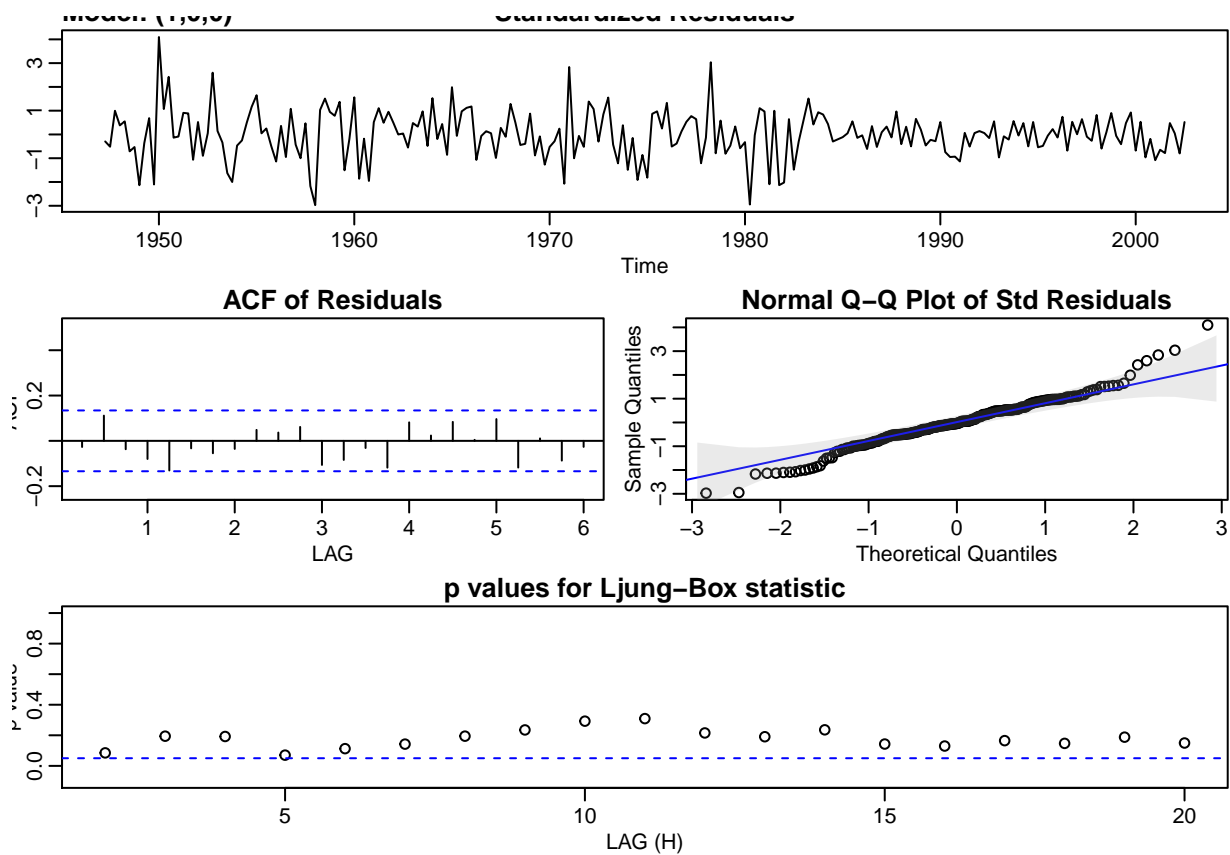
```
##
## Partial autocorrelations of series 'arma_sim3', by lag
##
##      1      2      3      4      5      6      7      8      9     10
## 0.949 -0.496  0.304 -0.222  0.074 -0.120  0.103 -0.109 -0.026 -0.030
##     11     12     13     14     15     16     17     18     19     20
## 0.028 -0.026 -0.018 -0.063 -0.001 -0.067 -0.007  0.002  0.006  0.001
##     21     22     23     24     25     26
## 0.006  0.007 -0.003  0.024 -0.022  0.016
```

```
ajust3 <- arima(arma_sim3, order = c(1,0,1), seasonal = c(0,0,0))
```

Exercício 31

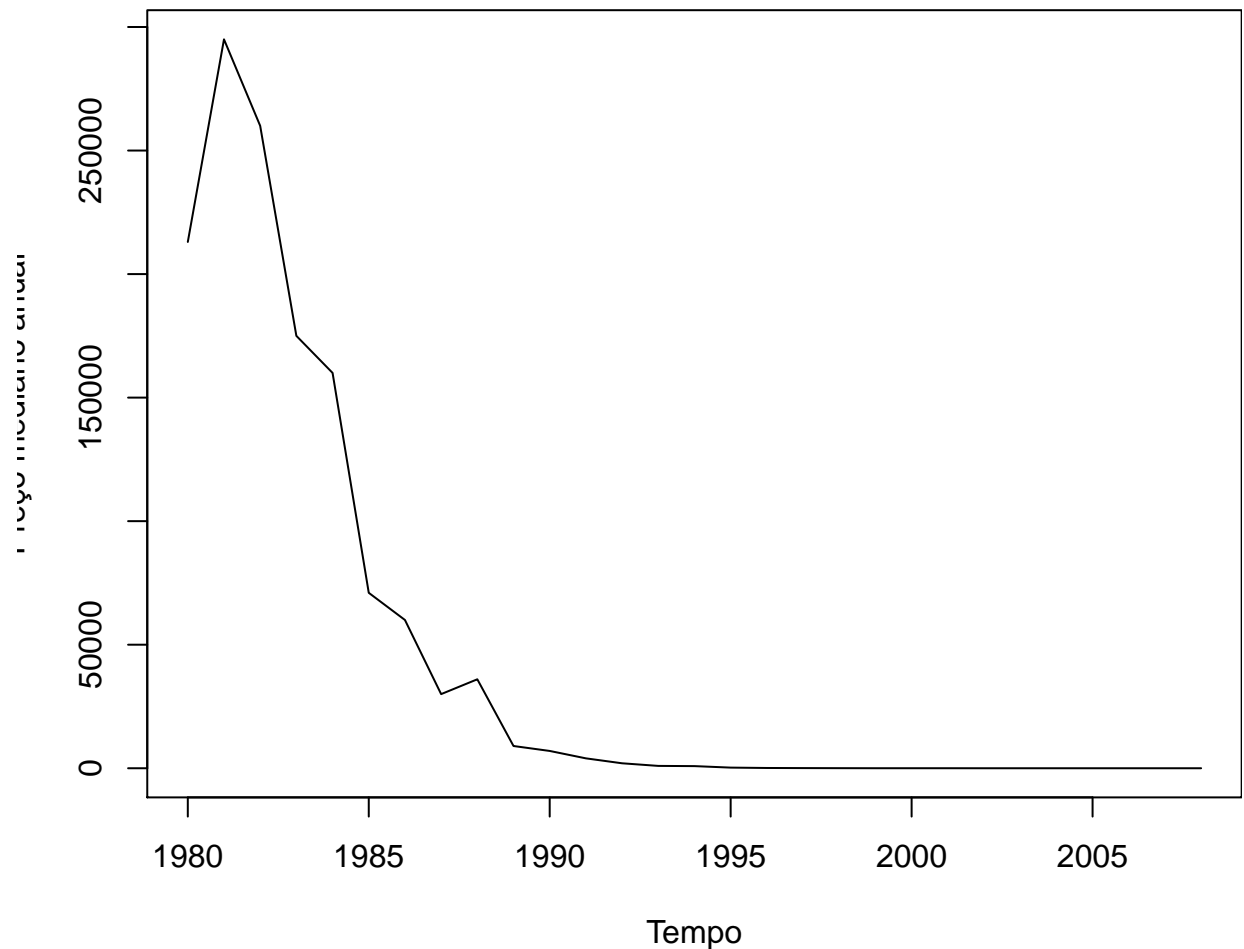
- I) Gráfico de resíduos padronizados: semelhante ao exemplo III.40, não há padrões claros no gráfico. Porém, devido a presença de valores discrepantes, algumas observações apresentam valores acima de três desvios padrões.
- II) ACF dos resíduos: não há nenhuma presença de auto-correlação evidente.
- III) Q-Qplot dos resíduos padronizados: há presença de valores discrepantes nas caudas. As observações entre -1 e 2 quantis teóricos estão bem ajustadas.
- IV) p-valores da estatística Q de Ljung-Box: nenhum p-valor abaixo do nível de significância, evidenciando a independência entre os resíduos.

```
gnpgr <- diff(log(gnp))
(Ex31 <- sarima(gnpgr,1,0,0))
```



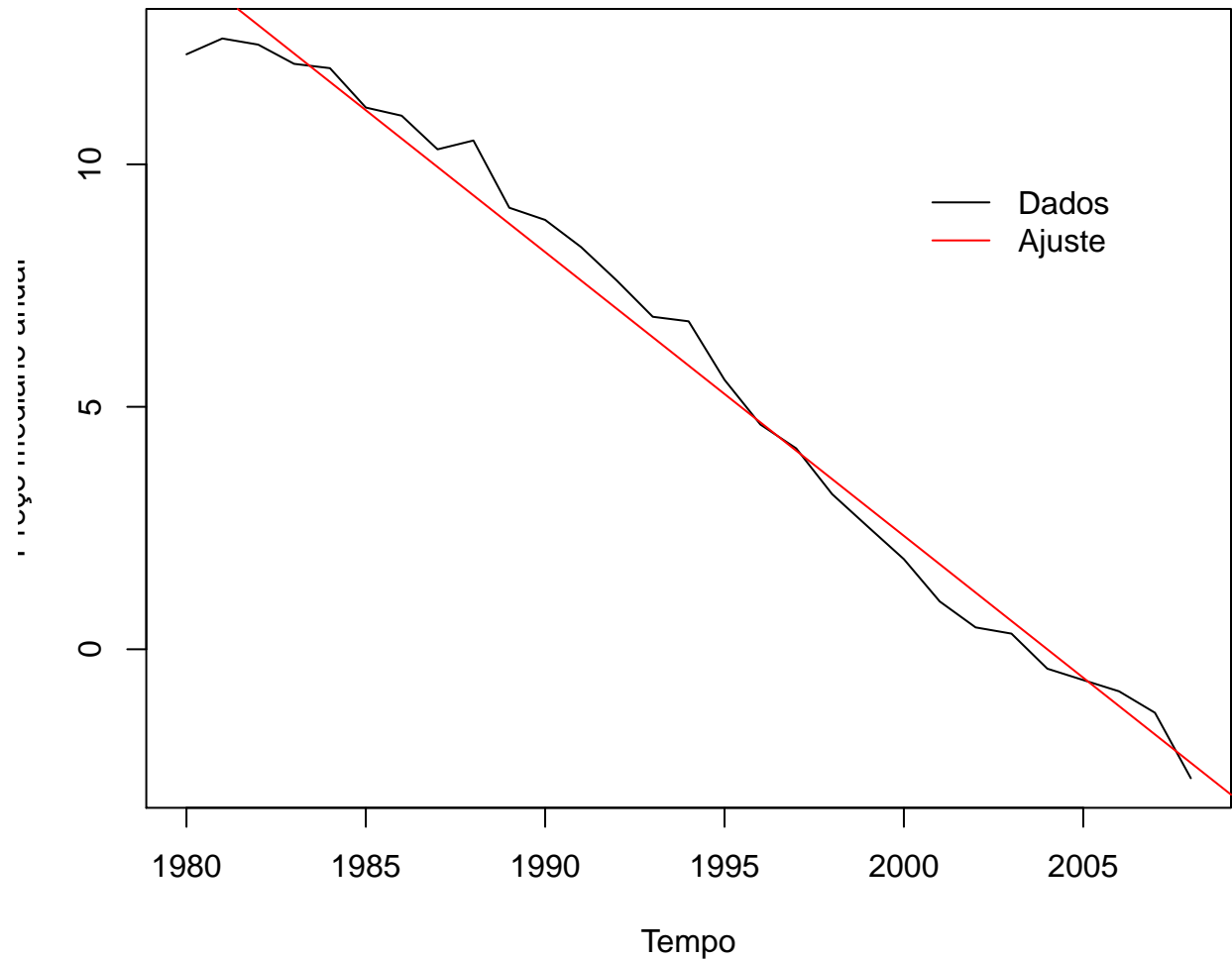
Exercício 36

- a) O comportamento da série temporal referente a variável preço mediano anual de varejo por GB de discos rígidos está apresentado na figura a seguir. De modo geral, é possível notar que a série não apresenta um comportamento estável ao longo do tempo. Inicialmente, há um aumento no valor da variável até 1983 e, na sequência, há uma evidente tendência de decrescimento da variável ao longo do tempo. Porém, a taxa de mudança nos valores da variável entre anos consecutivos é pequena a partir de 1988. Logo, os valores da variável se mostram bastante estáveis após 1990.



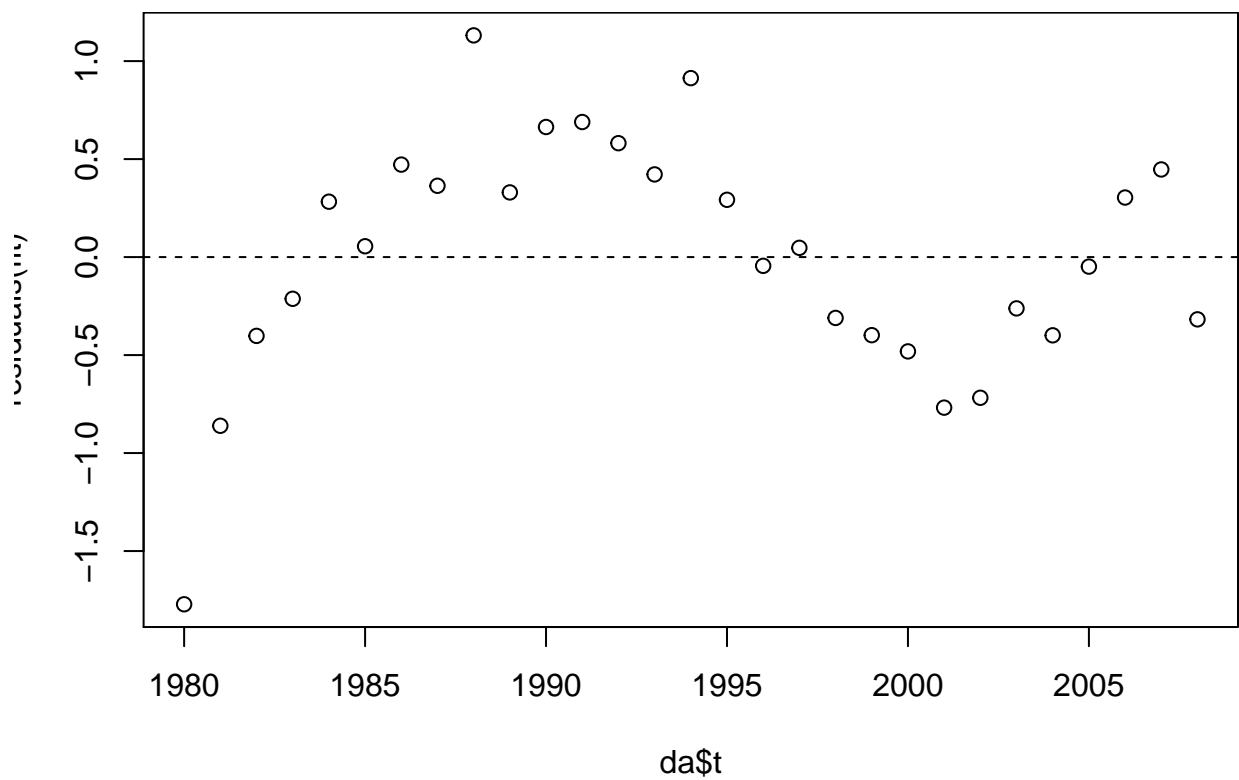
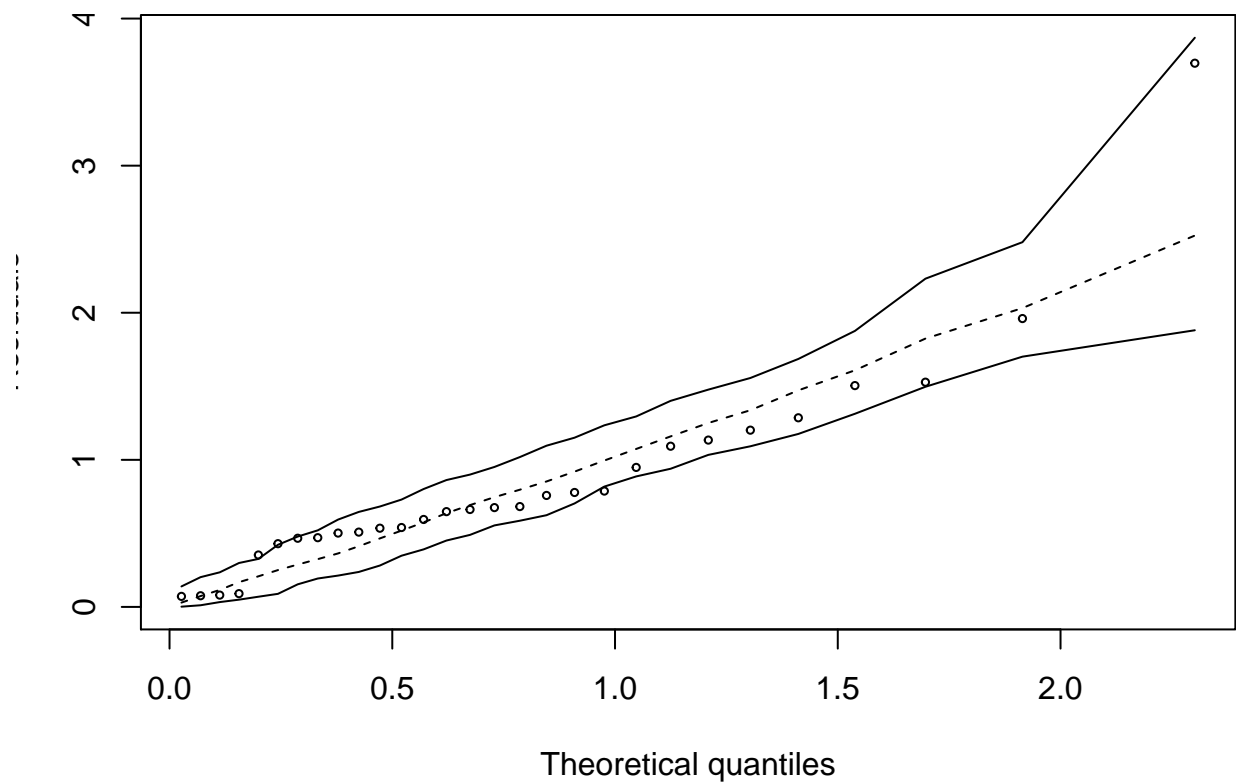
- b) O ajuste do modelo de regressão linear com a variável resposta transformada pelo logaritmo está apresentada na sequência. É possível notar que a transformação aplicada aos dados linearizou a relação da variável resposta e o tempo. Assim, o modelo de regressão linear apresentou um ajuste satisfatório aos dados e com elevado valor de coeficiente de determinação ajustado.

```
##
## Call:
## lm(formula = I(log(cpg)) ~ t, data = da)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77156 -0.39840  0.04726  0.42186  1.13129
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1172.49431   27.57793   42.52  <2e-16 ***
## t           -0.58508    0.01383  -42.30  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6231 on 27 degrees of freedom
## Multiple R-squared:  0.9851, Adjusted R-squared:  0.9846
## F-statistic: 1790 on 1 and 27 DF, p-value: < 2.2e-16
```



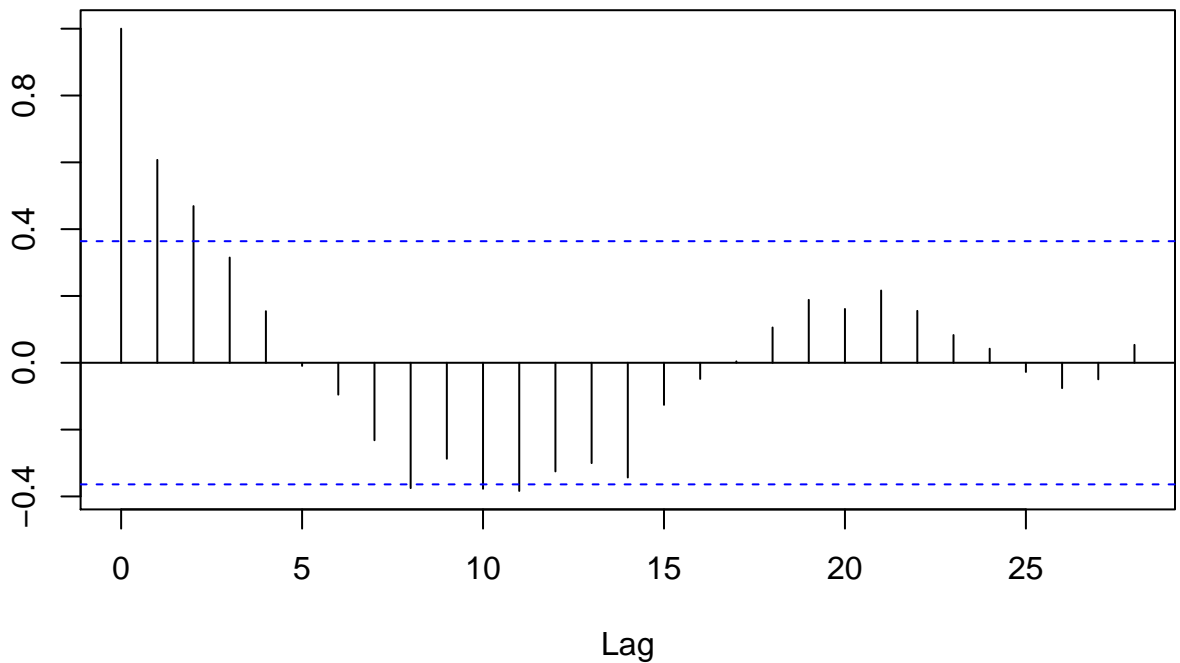
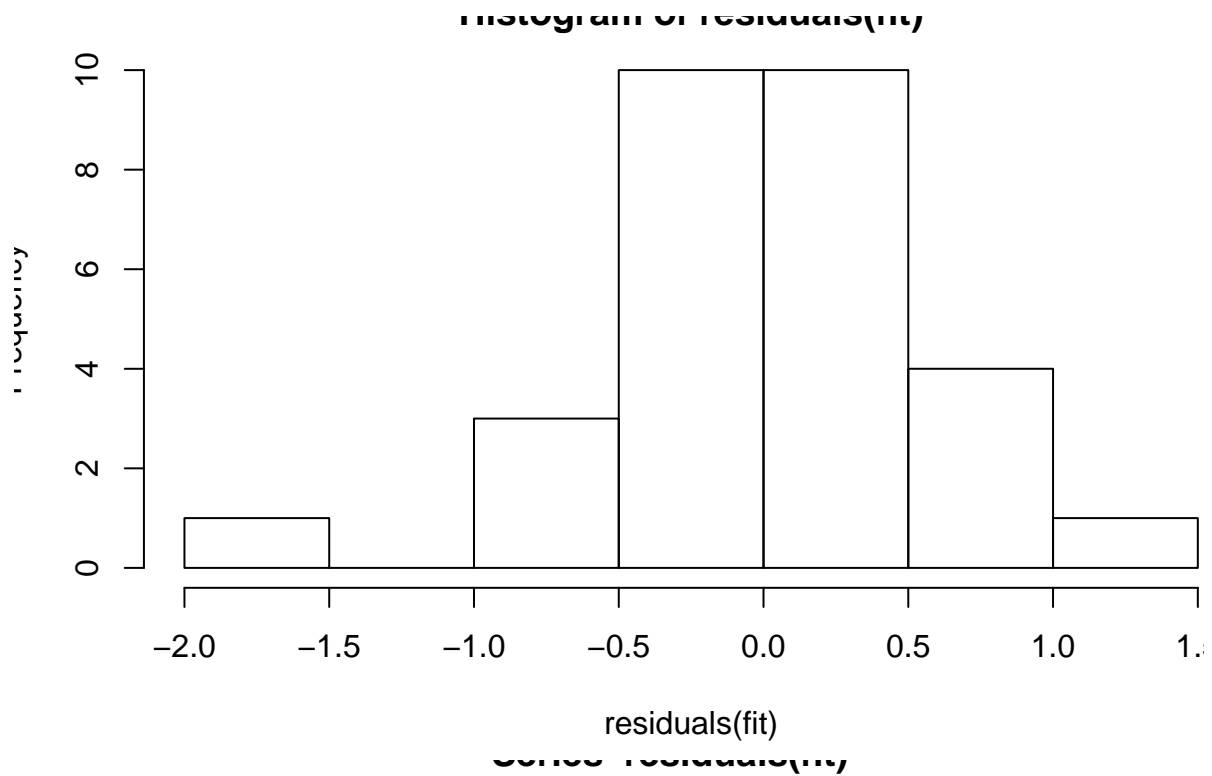
- c) O gráfico Half-Normal plot sugere que o modelo está bem especificado e ajustado aos dados, uma vez que há apenas uma observação fora dos envelopes simulados. Porém, o gráfico de dispersão dos resíduos ordinários em relação ao tempo mostra uma variância não constante ao longo do tempo, com um padrão cúbico dos resíduos. Esse resultado sugere a possibilidade de incluir a covariável tempo com seu efeito de terceiro grau, isto é, o tempo pode ser inserido com uma especificação polinomial de grau três no preditor linear. O histograma dos resíduos indicou normalidade, a qual foi confirmada pelo teste de normalidade de Shapiro-Wilk, em que a hipótese nula não foi rejeitada, para nível de 5% de significância. Por fim, o gráfico de autocorrelação dos resíduos indicou que há baixa correlação temporal.

Gaussian model (lm object)



```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit)
```

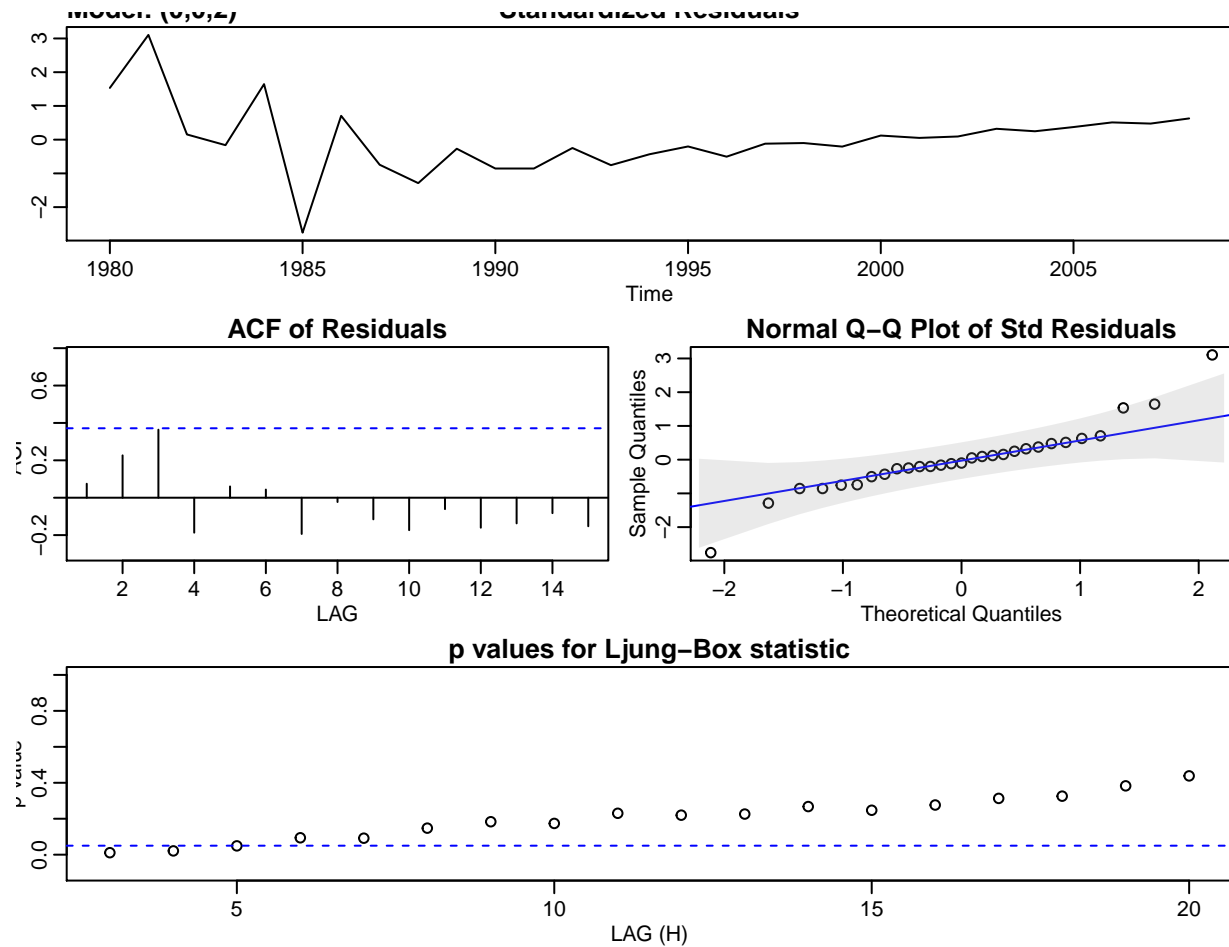
W = 0.96464, p-value = 0.4251



- d) O modelo de regressão ajustado para erros autocorrelacionados foi o SARIMA(0,0,2), isto é, o modelo ajustado foi um SARIMA com médias móveis de ordem dois. A figura apresentada na sequência mostra que os resíduos padronizados estão estáveis a partir do ano de 1990 entre anos consecutivos. Ainda, os resíduos do modelo ajustado se mostram não correlacionados. Também é possível notar que a distribuição dos resíduos se aproxima de uma distribuição normal, com

apenas dois valores extremos no gráfico qqplot. Por fim, é possível observar que o p-valor do teste Ljung-Box aumenta em função da defasagem, sugerindo que a hipótese nula de que os resíduos são independentes e identicamente distribuídos não é rejeitada para grande valores de defasagem.

```
## initial value 10.952140
## iter 2 value 10.383886
## iter 3 value 10.359737
## iter 4 value 10.346885
## iter 5 value 10.342800
## iter 6 value 10.341432
## iter 7 value 10.339494
## iter 8 value 10.339453
## iter 9 value 10.339440
## iter 10 value 10.339440
## iter 10 value 10.339439
## final value 10.339439
## converged
## initial value 10.344527
## iter 2 value 10.342504
## iter 3 value 10.330557
## iter 4 value 10.329140
## iter 5 value 10.328806
## iter 6 value 10.328739
## iter 7 value 10.328733
## iter 7 value 10.328733
## iter 7 value 10.328733
## final value 10.328733
## converged
```



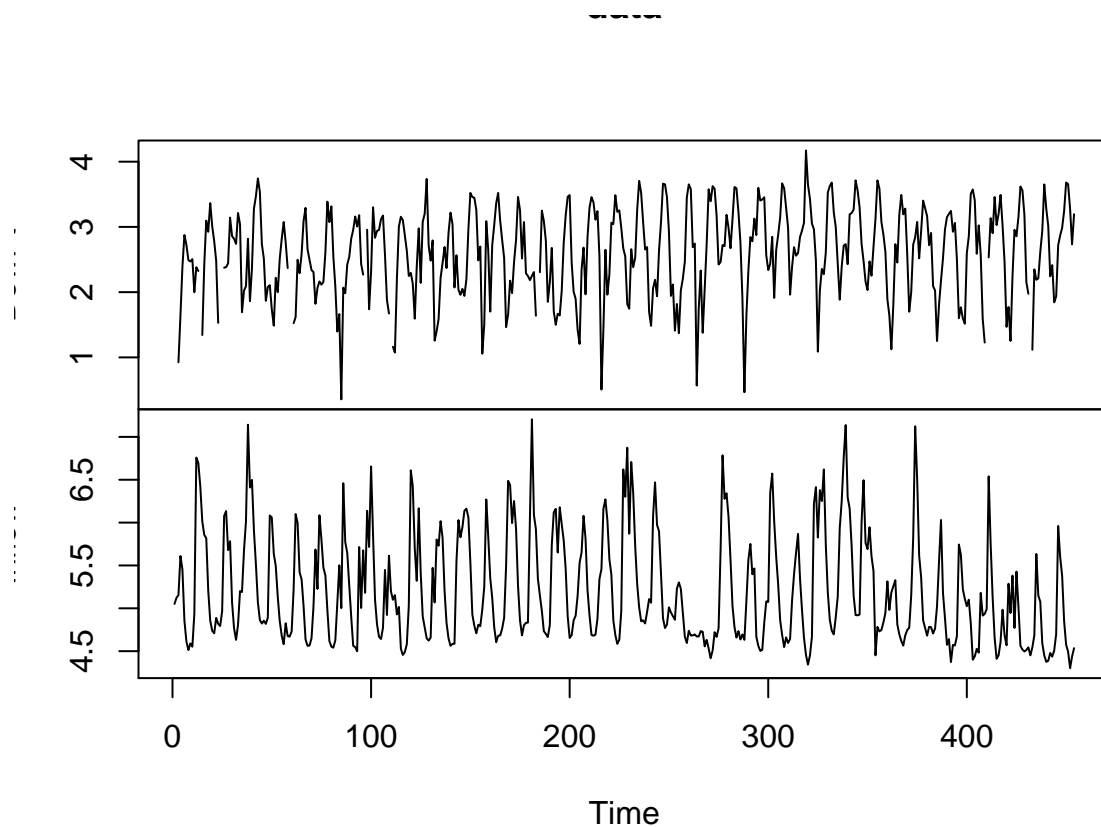
```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##      ma1      ma2  intercept      xreg
##      1.1476  0.7952  13017176 -6505.744
## s.e.  0.1864  0.2297   3542405  1776.370
##
## sigma^2 estimated as 858241410:  log likelihood = -340.68,  aic = 691.36
##
## $degrees_of_freedom
## [1] 25
##
## $ttable
##           Estimate      SE t.value p.value
## ma1           1.1476    0.1864  6.1558  0.0000
## ma2            0.7952    0.2297  3.4618  0.0019
## intercept 13017176.3326 3542404.9251  3.6747  0.0011
## xreg        -6505.7441   1776.3696 -3.6624  0.0012
##
```

```
## $AIC
## [1] 23.84017
##
## $AICc
## [1] 23.89764
##
## $BIC
## [1] 24.07591
```

PARTE IV - Análise Espectral e Filtragem

Exercício 6a

- a)



```
##
## Call:
## arima(x = data[, 1], order = c(0, 0, 0), seasonal = list(order = c(0, 1, 1),
##   period = 12))
##
## Coefficients:
##      sma1
##    -0.8670
## s.e.   0.0285
##
## sigma^2 estimated as 0.1816:  log likelihood = -252.46,  aic = 508.91
```

```
##
## Call:
## arima(x = data[, 2], order = c(0, 0, 0), seasonal = list(order = c(0, 1, 1),
##   period = 12))
##
## Coefficients:
##          sma1
##        -0.8878
## s.e.    0.0330
##
## sigma^2 estimated as 0.1998:  log likelihood = -280.57,  aic = 565.13
```

Exercício 6b

- b) Os resultados encontrados sugerem três tipos de comportamento para as séries analisadas. Inicialmente, há correlação cruzada negativa, a qual vai diminuindo até não haver correlação. Posteriormente, há correlação cruzada positiva, seguido de correlação cruzada negativa. Os comportamentos distintos podem ser atribuídos a sazonalidade dos dados.

