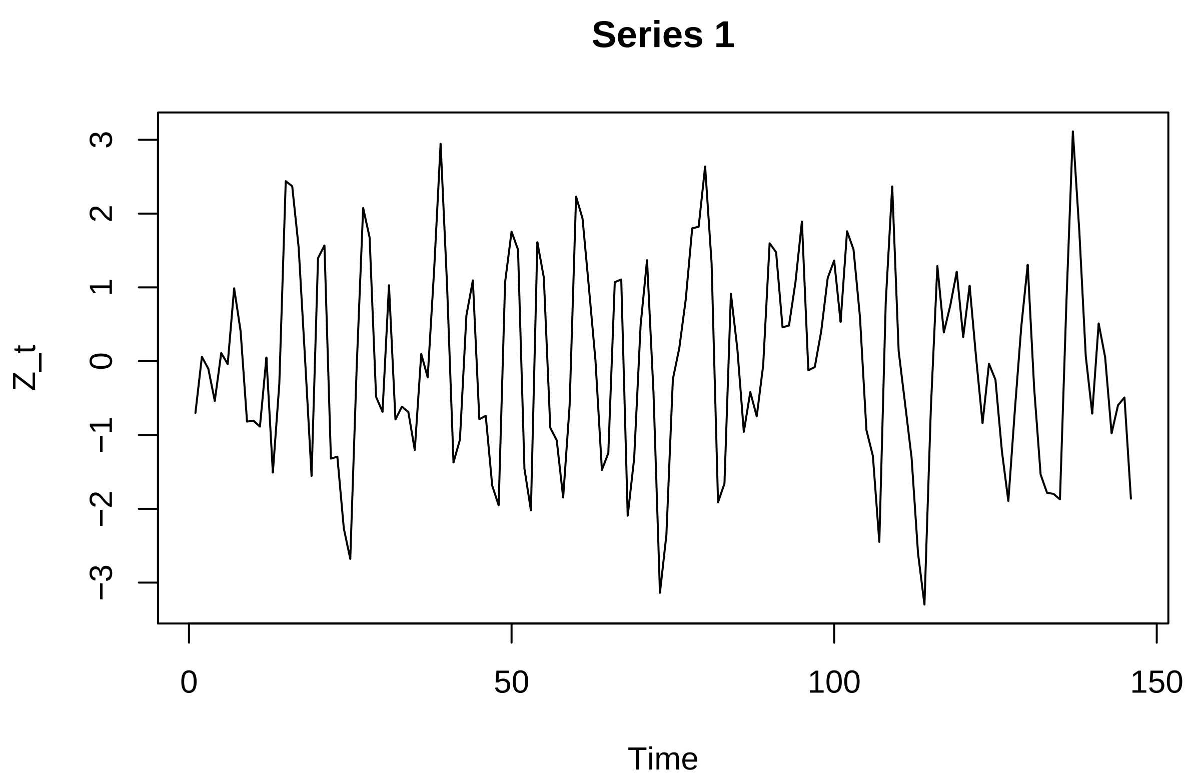
Modelos Arma

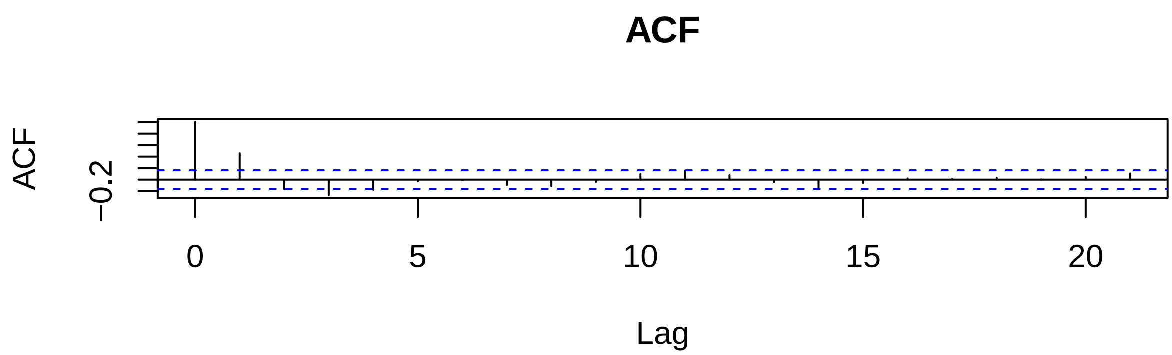
Giovani Valdrighi, Vitória Guardieiro

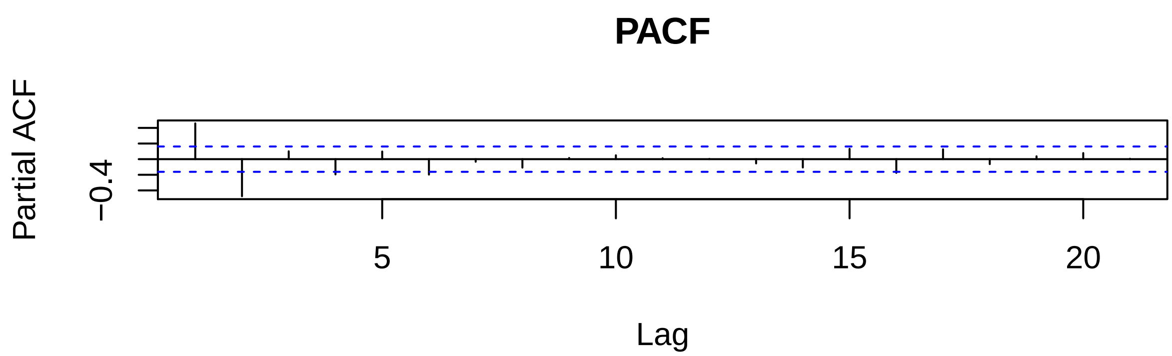
Modelos ARMA

Com 9 séries temporais, iremos avaliar cada uma delas e identificar se ela é gerada por um modelo AR(p), um modelo MA(q) ou um modelo ARMA(p, q). Em todas as diferentes séreis iremos inicialmente visualizar a série, a função de autocorrelação e a função de autocorrelação parcial.

Série 1



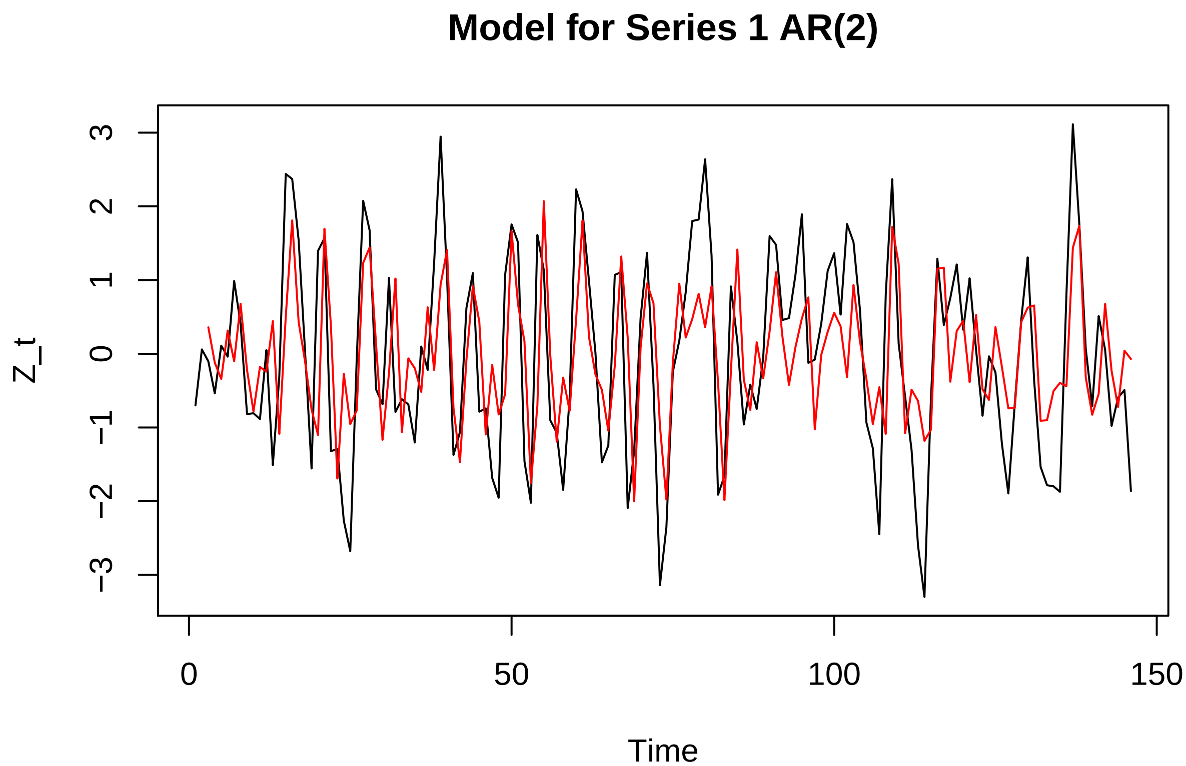




Vemos que a ACF e a PACF decrescem de forma brusca, tendo valores significativos para os dois primeiros lags, dessa forma, iremos considerar dois modelos, o modelo AR(2) e o modelo ARMA(2, 2). Vamos tentar fitar um modelo AR(2).

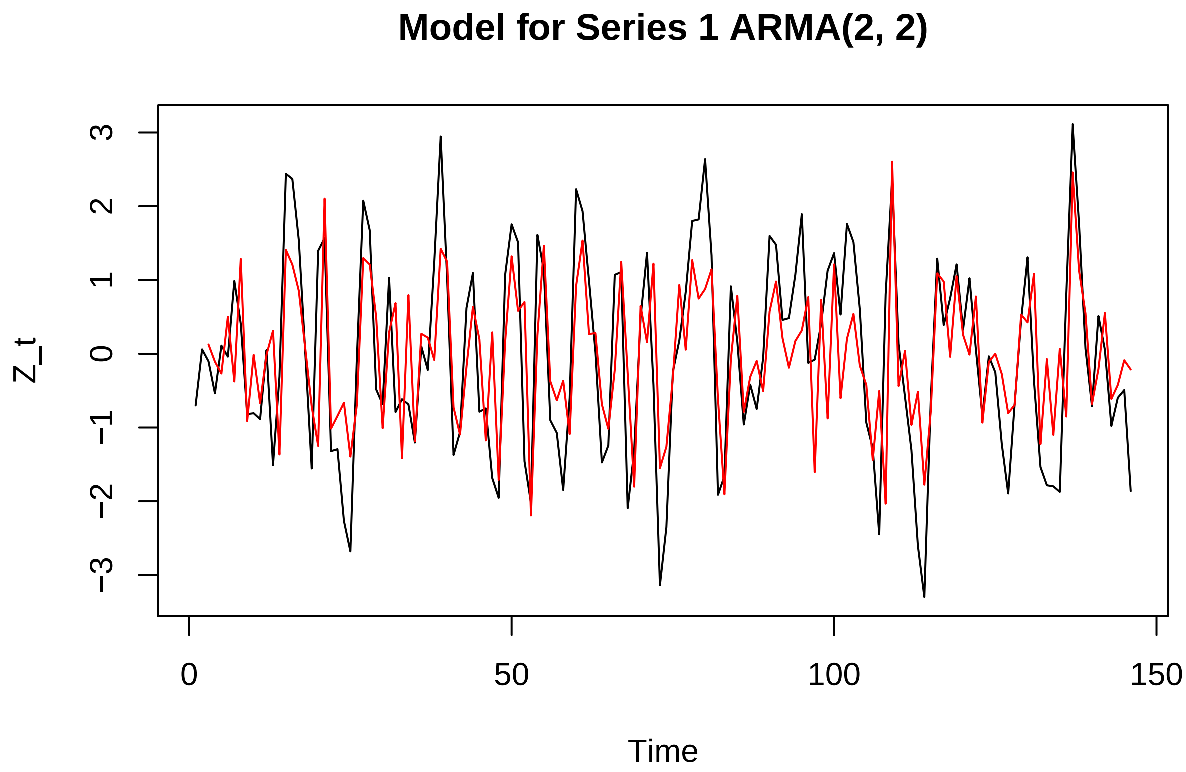
```
##
## Call:
## arma(x = X[[1]], order = c(2, 0))
##
## Model:
## ARMA(2,0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3168 -0.7743  0.1256  0.6870  2.4984
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ar1         0.68921     0.07292    9.451  < 2e-16 ***
## ar2        -0.48531     0.07289   -6.658 2.77e-11 ***
## intercept  -0.02129     0.08702   -0.245    0.807
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 1.113,  Conditional Sum-of-Squares = 159.18,  AIC = 435.98
```

Vamos visualizar inicialmente o modelo real e o previsto, e em sequência, o plot de resíduos.

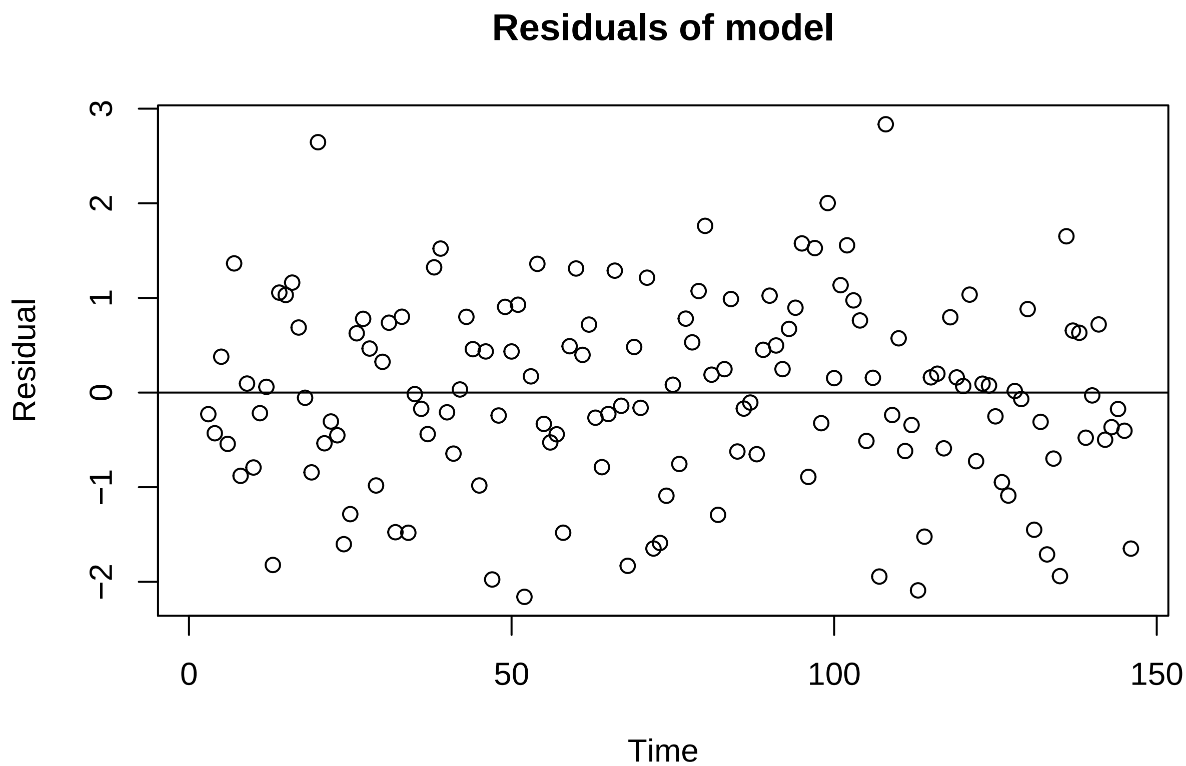


Realizando o mesmo procedimento, mas para o modelo ARMA(2, 2).

```
##
## Call:
## arma(x = X[[1]], order = c(2, 2))
##
## Model:
## ARMA(2,2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -2.1588765 -0.5967001  0.0001618  0.7185817  2.8350422
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ar1         0.792873    0.133253    5.950 2.68e-09 ***
## ar2        -0.119687    0.107696   -1.111    0.266
## ma1         0.060935    0.110778    0.550    0.582
## ma2        -0.775825    0.116414   -6.664 2.66e-11 ***
## intercept -0.005245    0.024207   -0.217    0.828
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 0.9758,  Conditional Sum-of-Squares = 139.55,  AIC = 420.76
```

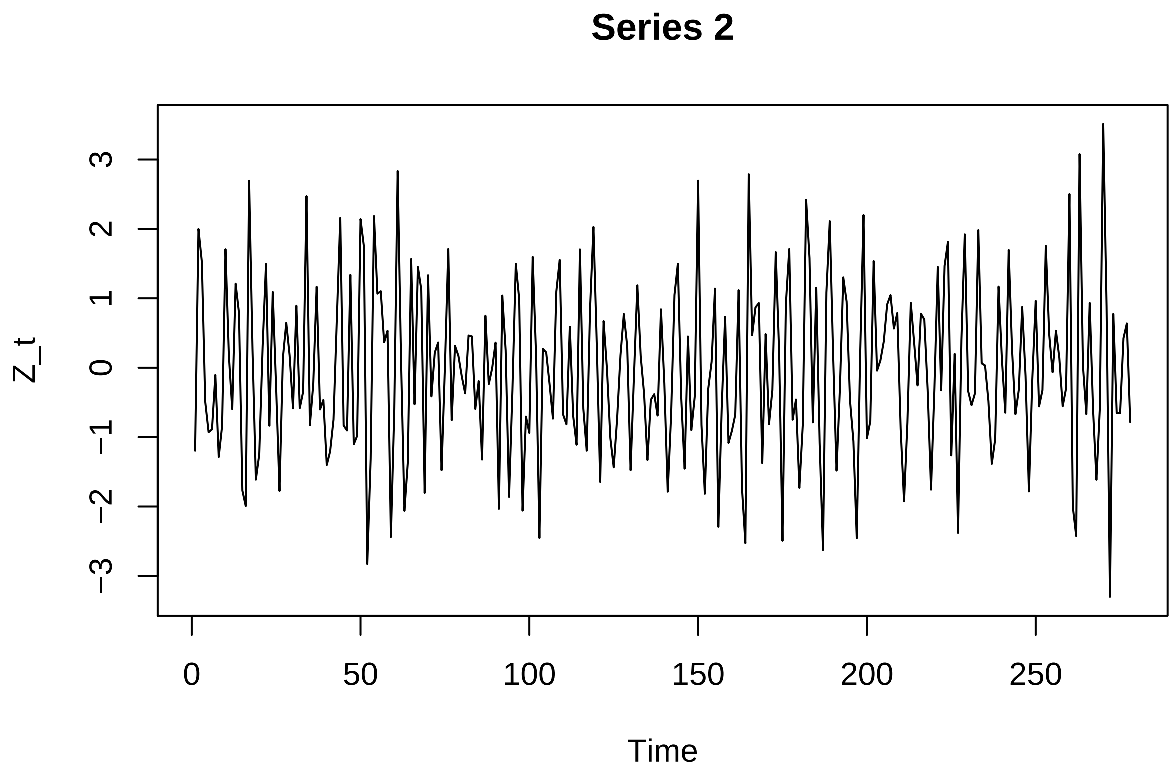


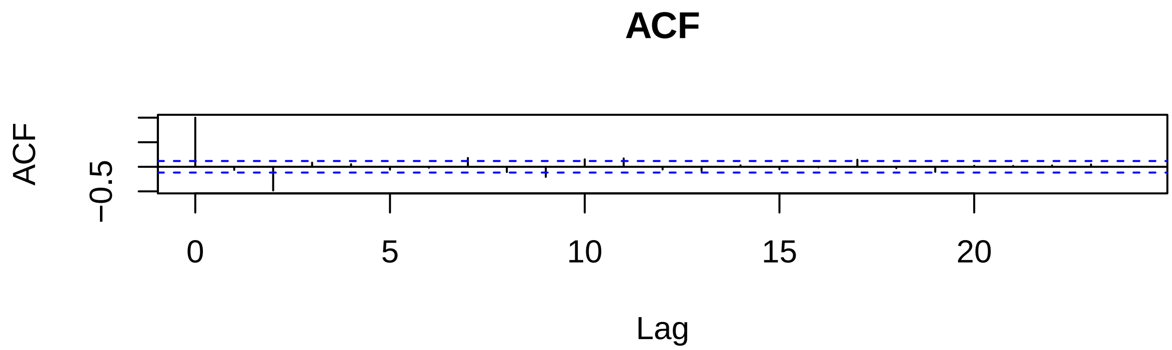
Vemos que o AIC é menor para o modelo ARMA(2, 2), e apesar de ambos se encaixarem bem aos dados, o modelo ARMA(2, 2) representa melhor os pontos extremos da série. Vamos visualizar os resíduos obtidos:

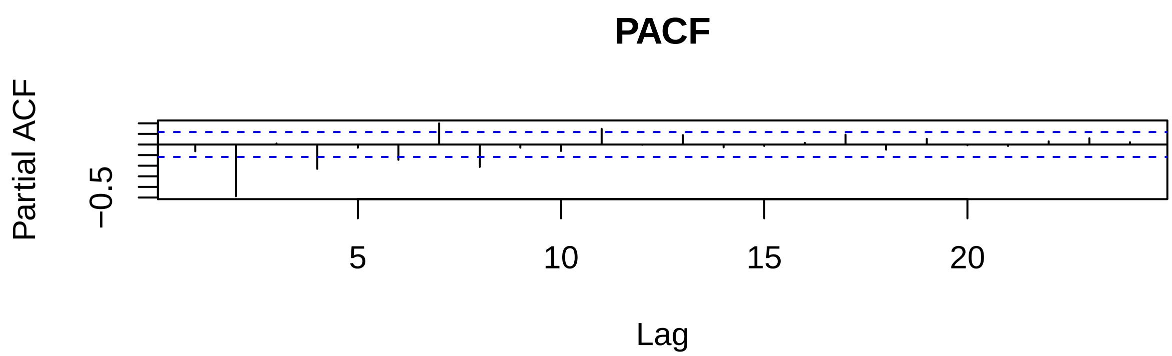


O modelo parece se adequar bem aos dados e também os resíduos não apresentam um padrão de comportamento.

Série 2





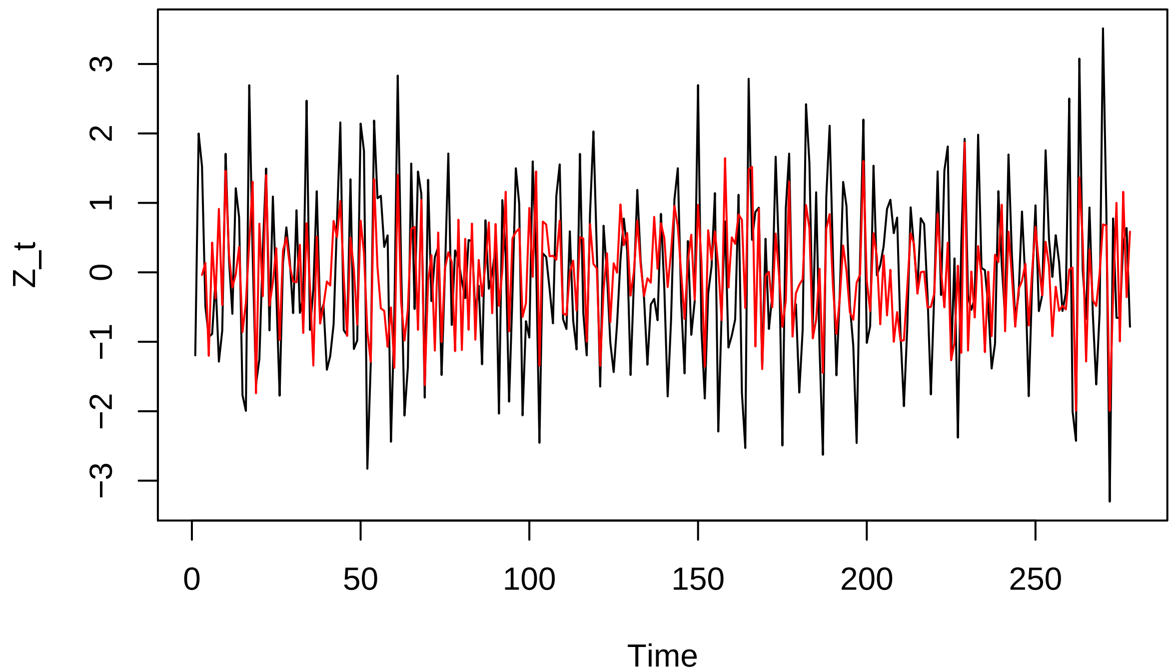


Agora, visualizamos um situação inversa, o lag 2 é significativo na ACF e nos demais não, e na PACF o decrescimento é gradual, o que nos faz pensar se tratar de um modelo MA(2). No entanto, como na PACF o lag 2 também é significativo, iremos comparar com o modelo ARMA(2, 2).

```
##
## Call:
## arma(x = X[[2]], order = c(0, 2))
##
## Model:
## ARMA(0,2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.49044 -0.77251  0.01282  0.77130  2.82319
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ma1         0.11057     0.04029    2.745  0.00606 **
## ma2        -0.70037     0.03931  -17.818  < 2e-16 ***
## intercept  -0.03815     0.02510   -1.520  0.12855
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 1.025,  Conditional Sum-of-Squares = 281.95,  AIC = 801.86
```

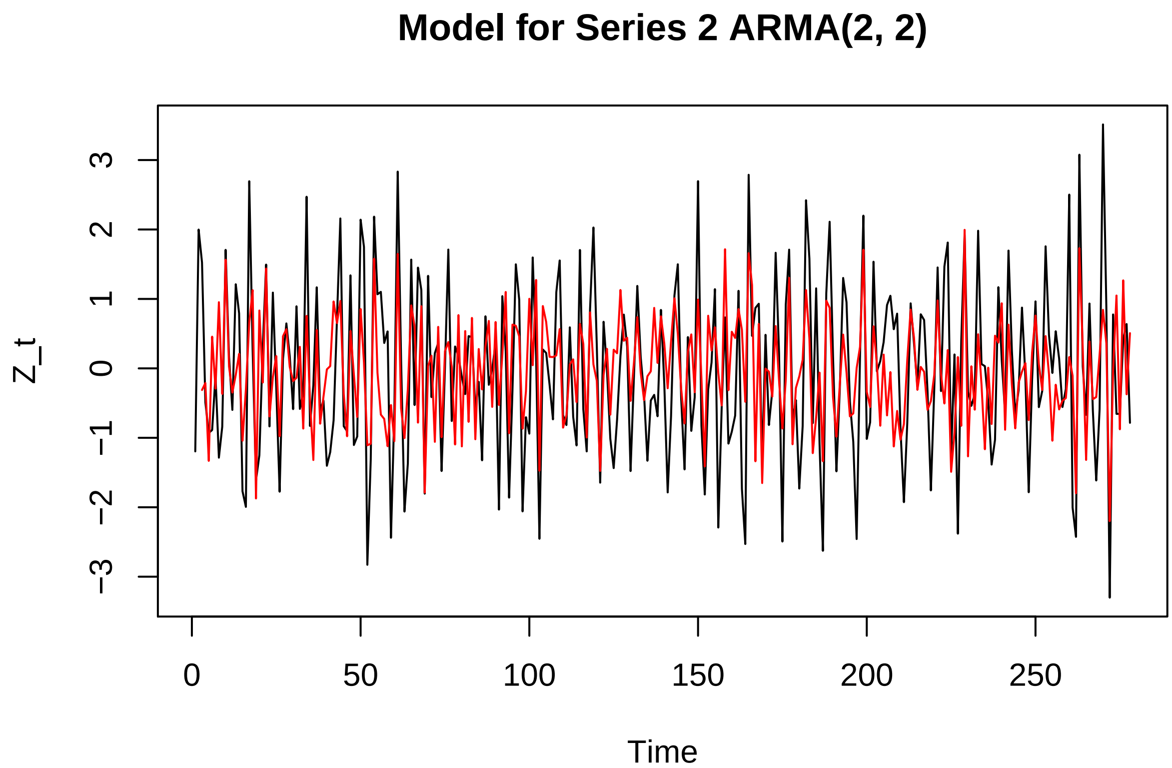
O modelo se encaixou bem, vamos comparar a previsão e o real, e em sequência, o plot de resíduos.

Model for Series 2 MA(2)

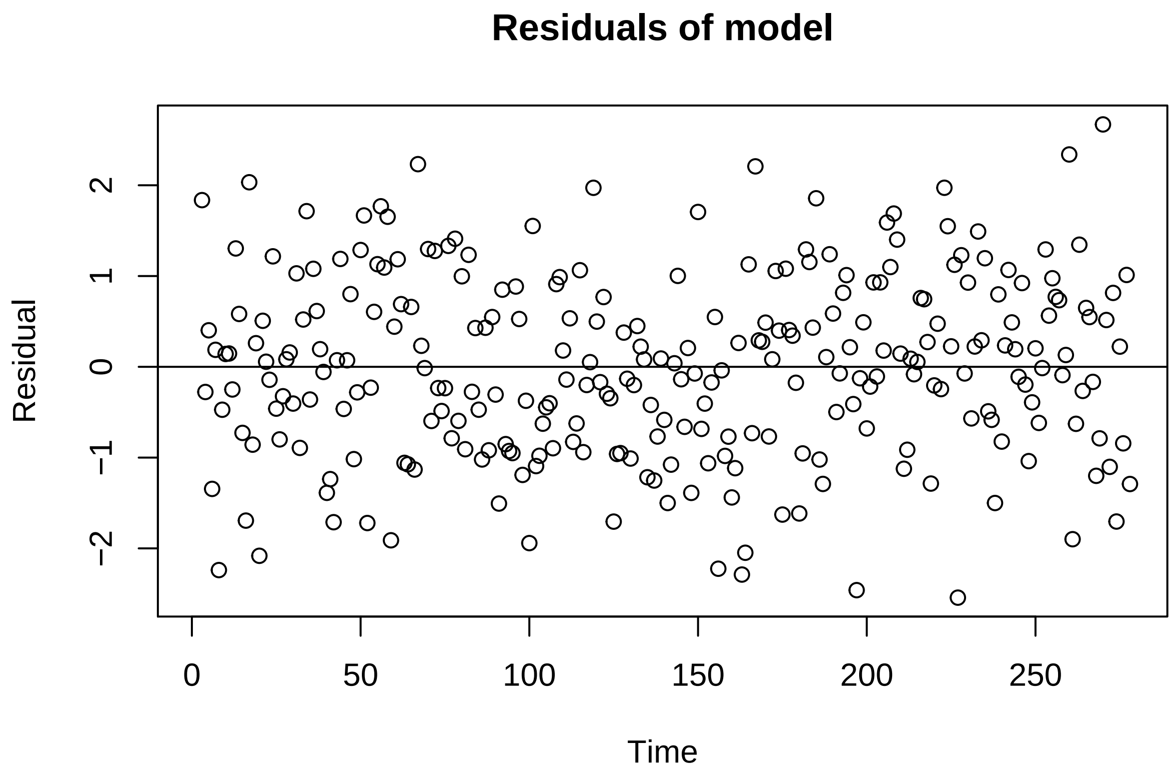


Realizando o mesmo processo com o modelo ARMA(2, 2).

```
##
## Call:
## arma(x = X[[2]], order = c(2, 2))
##
## Model:
## ARMA(2,2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.54256 -0.76754  0.01269  0.73690  2.67023
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ar1        -0.20365     0.08558   -2.380  0.01733 *
## ar2        -0.12123     0.07719   -1.571  0.11627
## ma1         0.21277     0.06716    3.168  0.00153 **
## ma2        -0.61986     0.06320   -9.808  < 2e-16 ***
## intercept  -0.04935     0.03605   -1.369  0.17106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 1.003,  Conditional Sum-of-Squares = 275.78,  AIC = 799.72
```

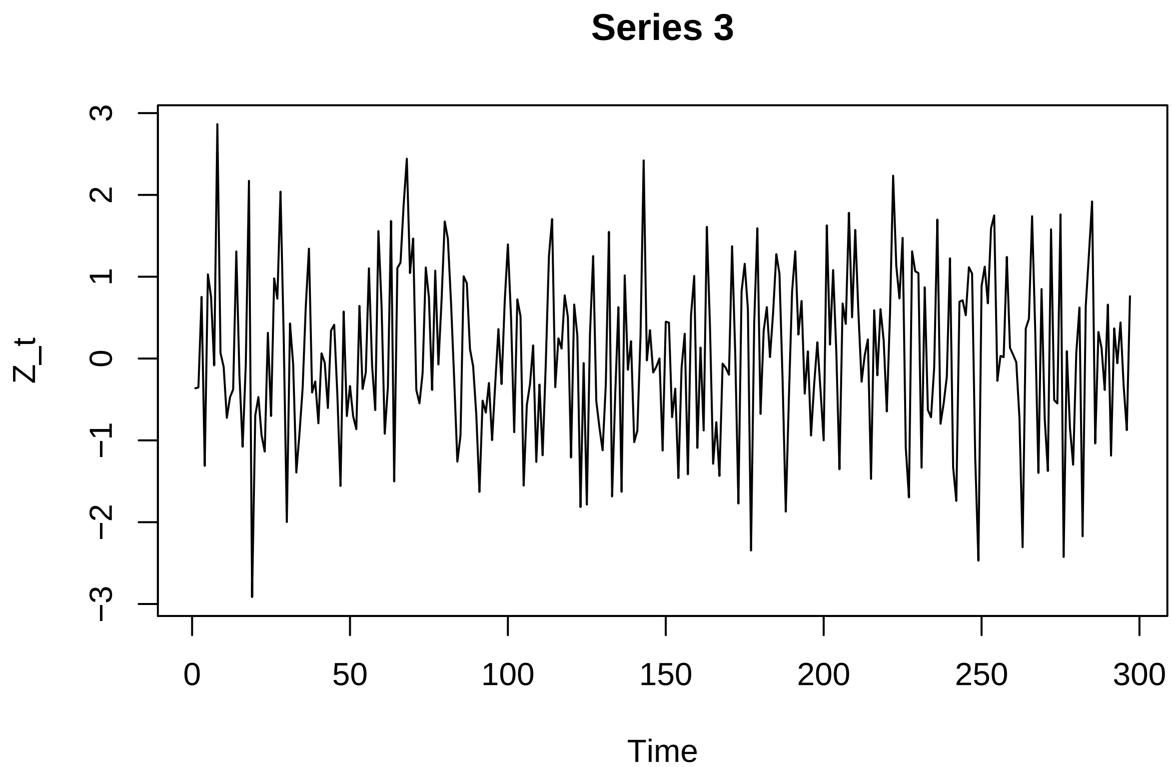


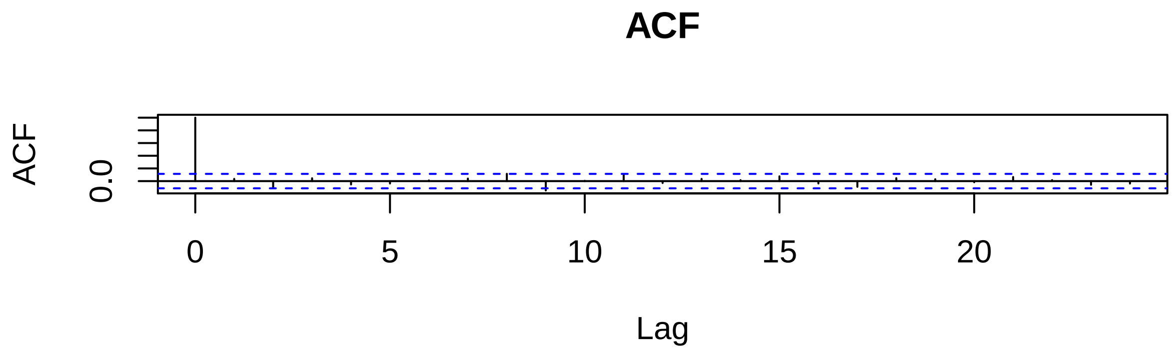
Vemos que apesar de extramemente parecidos, o modelo ARMA(2, 2) teve o AIC um pouquinho menor. Vamos visualizar o resíduo obtido com ele:

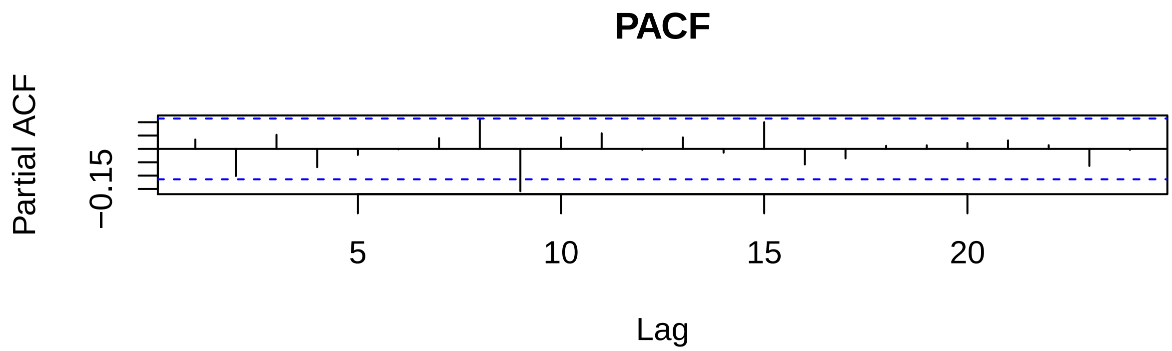


O modelo aparentemente se adequa bem a sazonalidade da série real, no entanto, não conseguimos capturar os picos extremos como ocorrem na série real, e os resíduos também se distribuem uniformemente ao longo da série.

Série 3



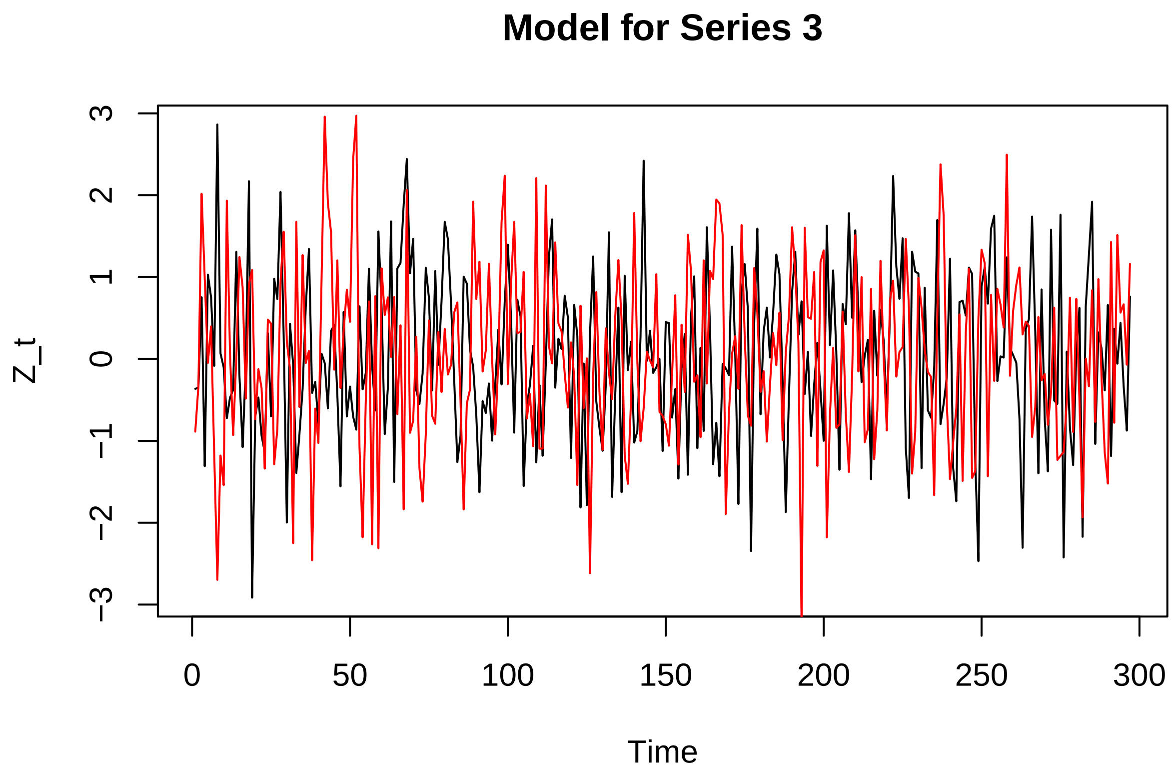




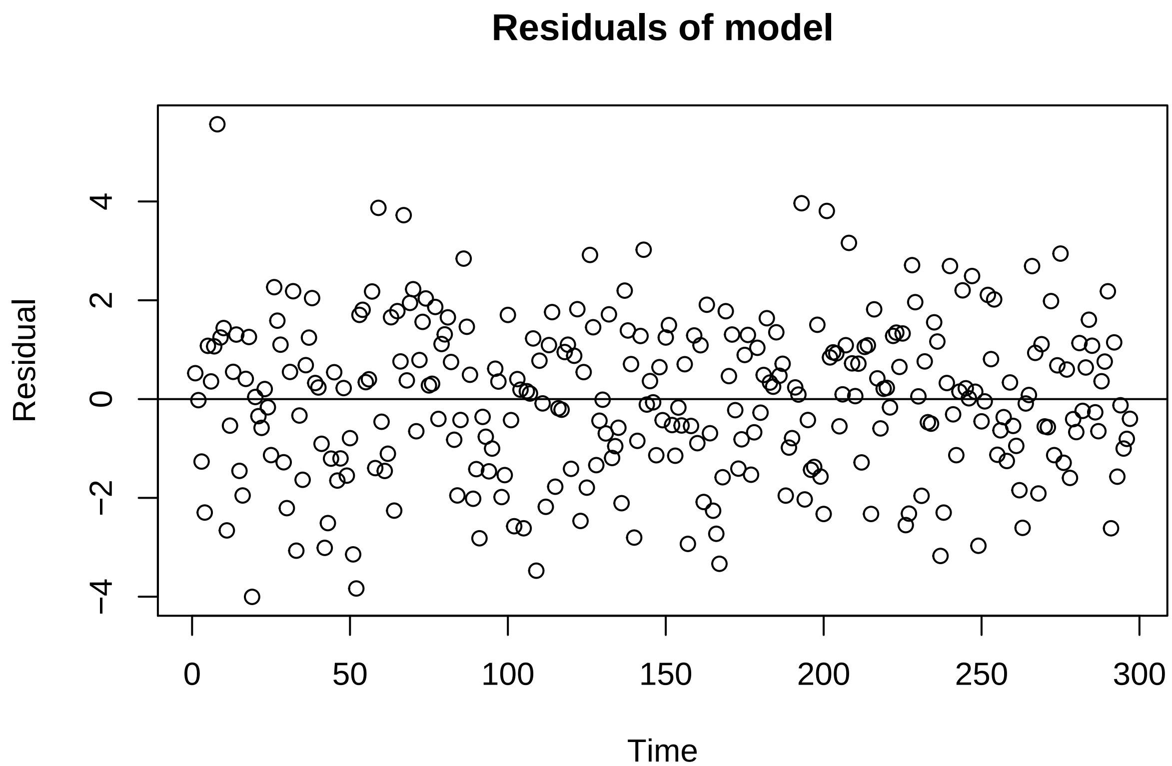
Nesse modelo existe um comportamento diferente dos demais, tanto a ACF quanto a PACF são praticamente nulas para todos os valores, menos para a ACF de 0, o que indica que as amostras não possuem covariância, se comportando como um ruído branco. Vamos verificar a média e a variância da série.

```
##          mean variance
## 1 0.03941307 1.021629
```

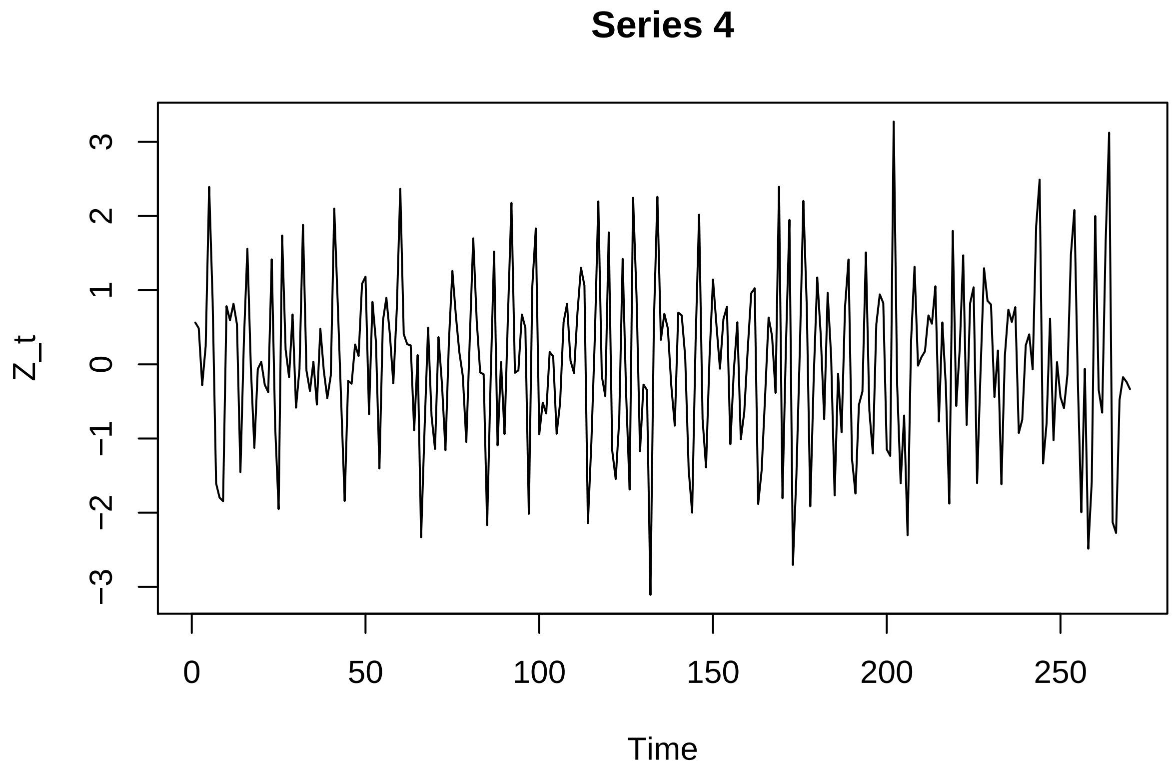
Vemos que o modelo se comporta como um ruído branco, isto é, a_t com $E(a_t) = 0$ e $Var(a_t) = 1$. Se nós gerarmos 297 amostras de a_t e visualizarmos tanto o modelo e predição, quanto o residual, teremos:



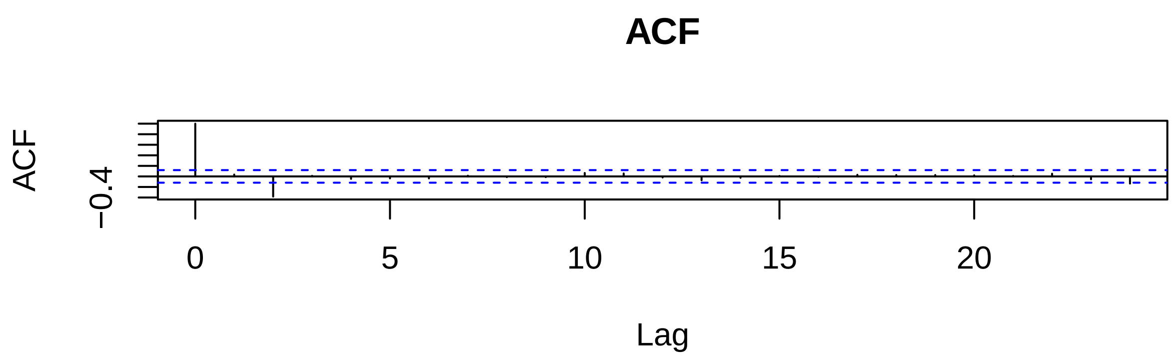

```
## Warning in model3$residual <- X[[3]] - model3: Realizando coerção de LHD para
## uma lista
```


Residuals of model
Residual
4
2
0
−2
−4
0
50
100
150
200
250
300
Time

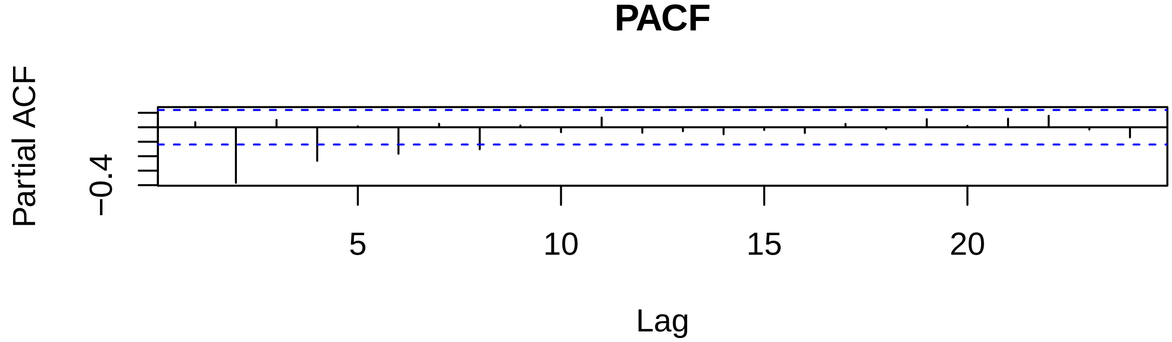
Série 4



Series 4
Z_t
3
2
1
0
−1
−2
−3
0
50
100
150
200
250
Time



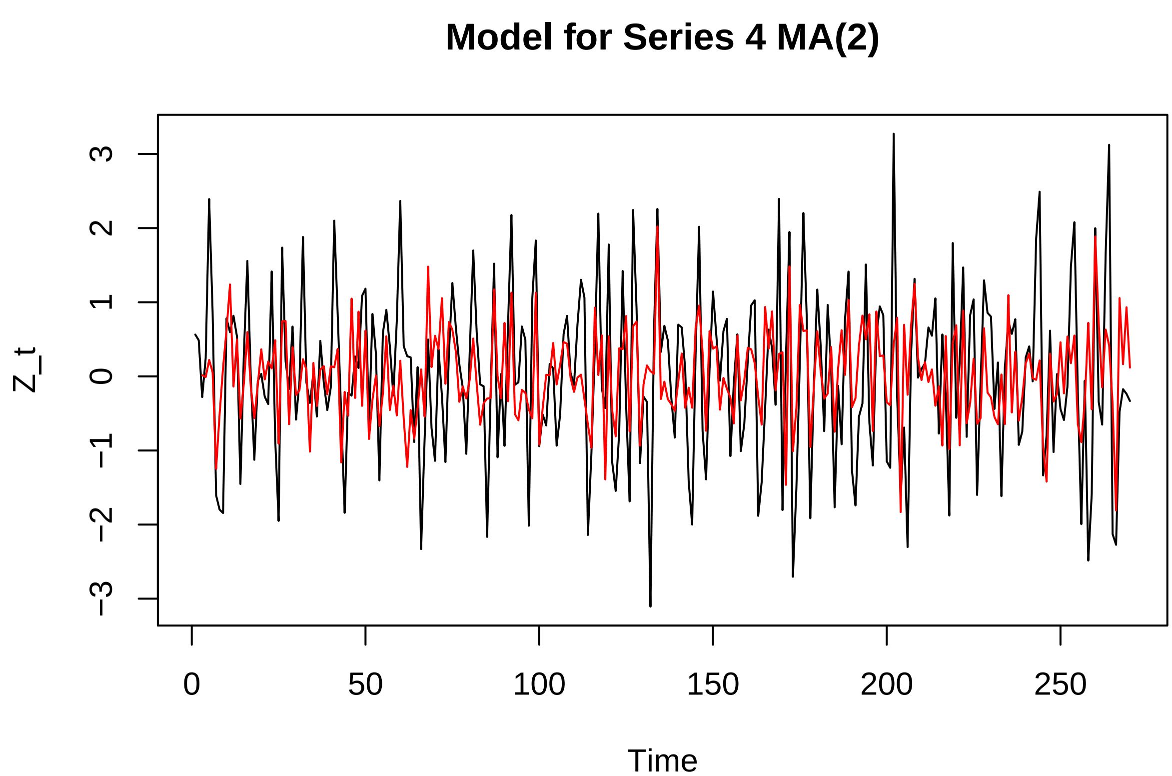
ACF
ACF
−0.4
0
5
10
15
20
Lag

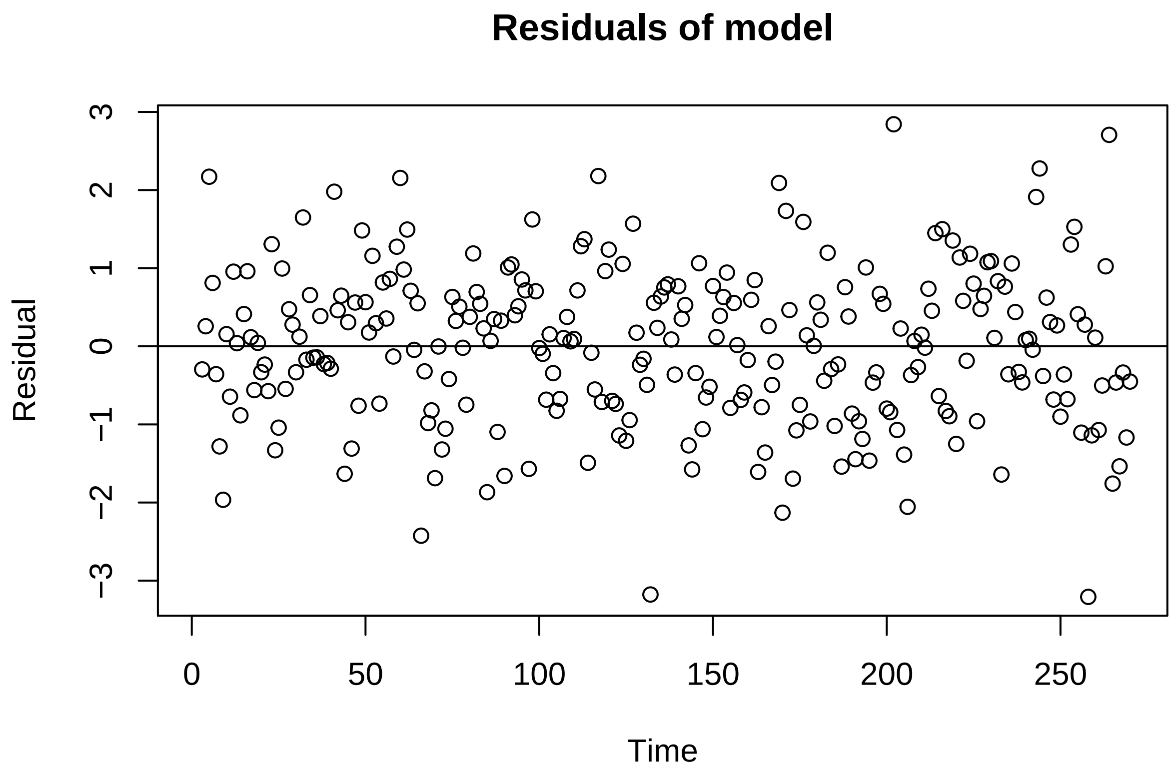


PACF
Partial ACF
−0.4
5
10
15
20
Lag

A série apresenta uma função de ACF que cai drasticamente após o lag 2, enquanto a PACF cai mais gradualmente de forma exponencial, o que dá a noção de se tratar de um modelo MA(2). Utilizando dessa observação, fitamos:

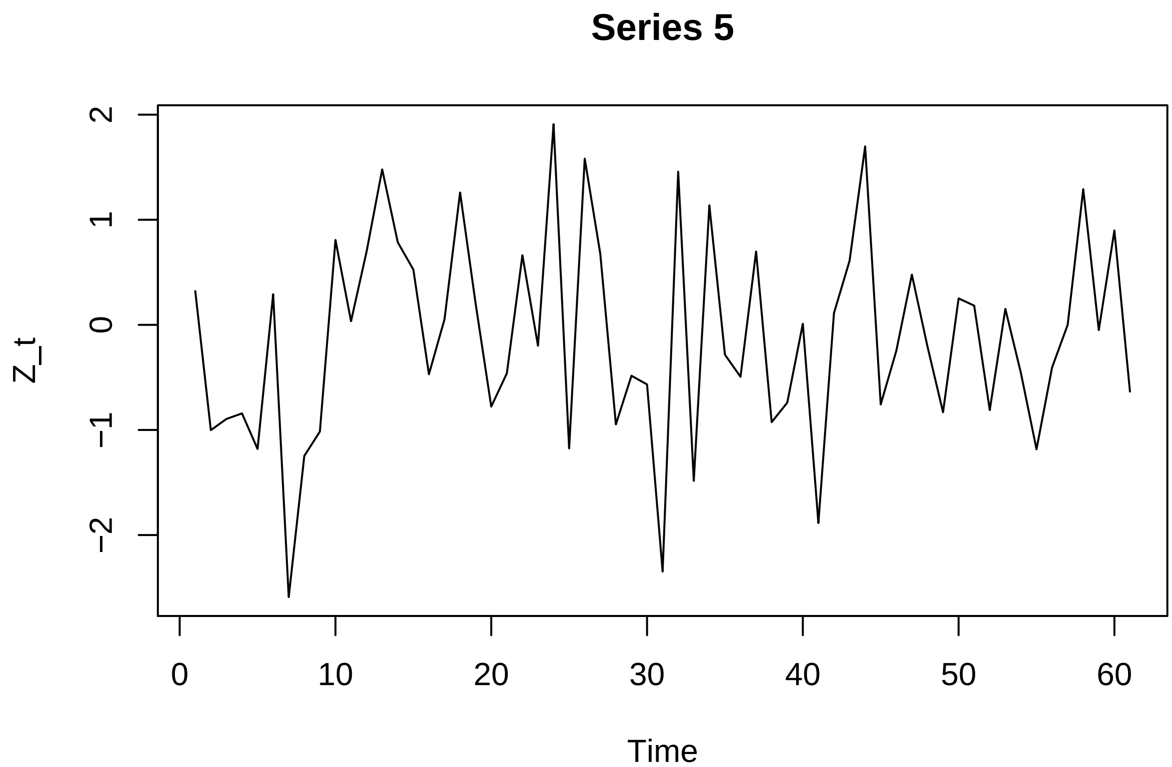
```
##
## Call:
## arma(x = X[[4]], order = c(0, 2))
##
## Model:
## ARMA(0,2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.20748 -0.68404  0.06891  0.64994  2.84235
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ma1         0.09013     0.05112    1.763   0.0779 .
## ma2        -0.61589     0.05318  -11.581   <2e-16 ***
## intercept   0.01590     0.02884    0.551   0.5813
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 0.9865,  Conditional Sum-of-Squares = 263.41,  AIC = 768.57
```

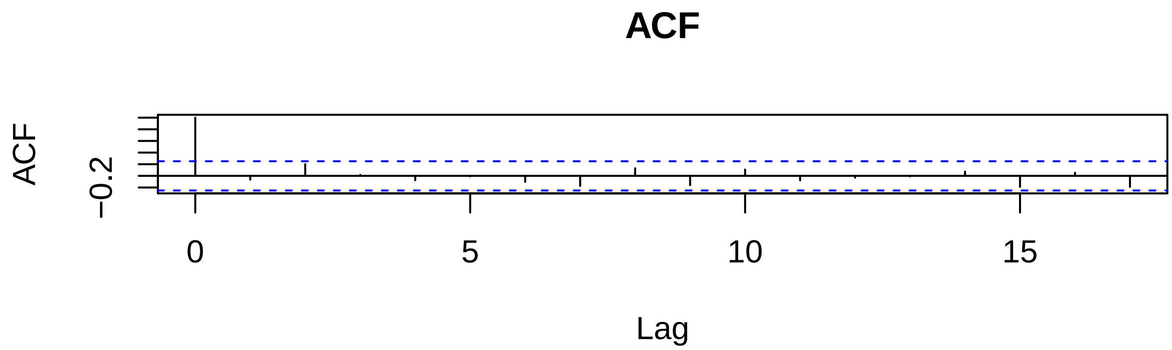


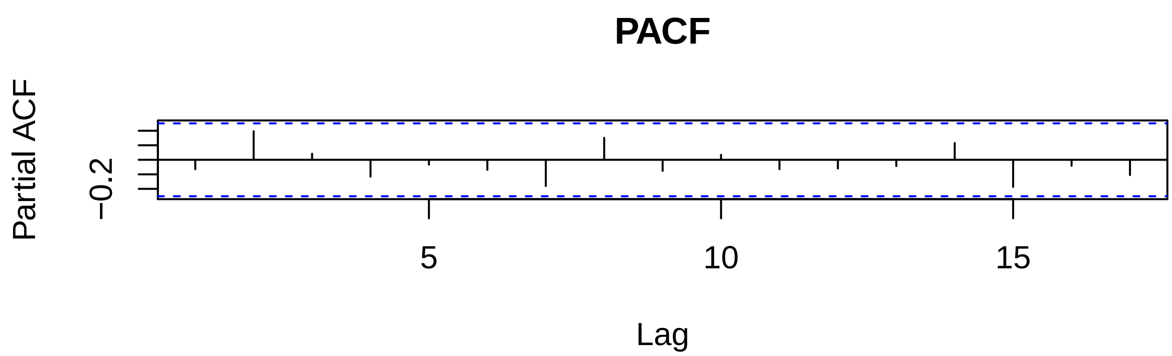


O modelo se encaixou bem aos dados, incluindo apresentando uma distribuição uniforme dos ruídos.

Série 5

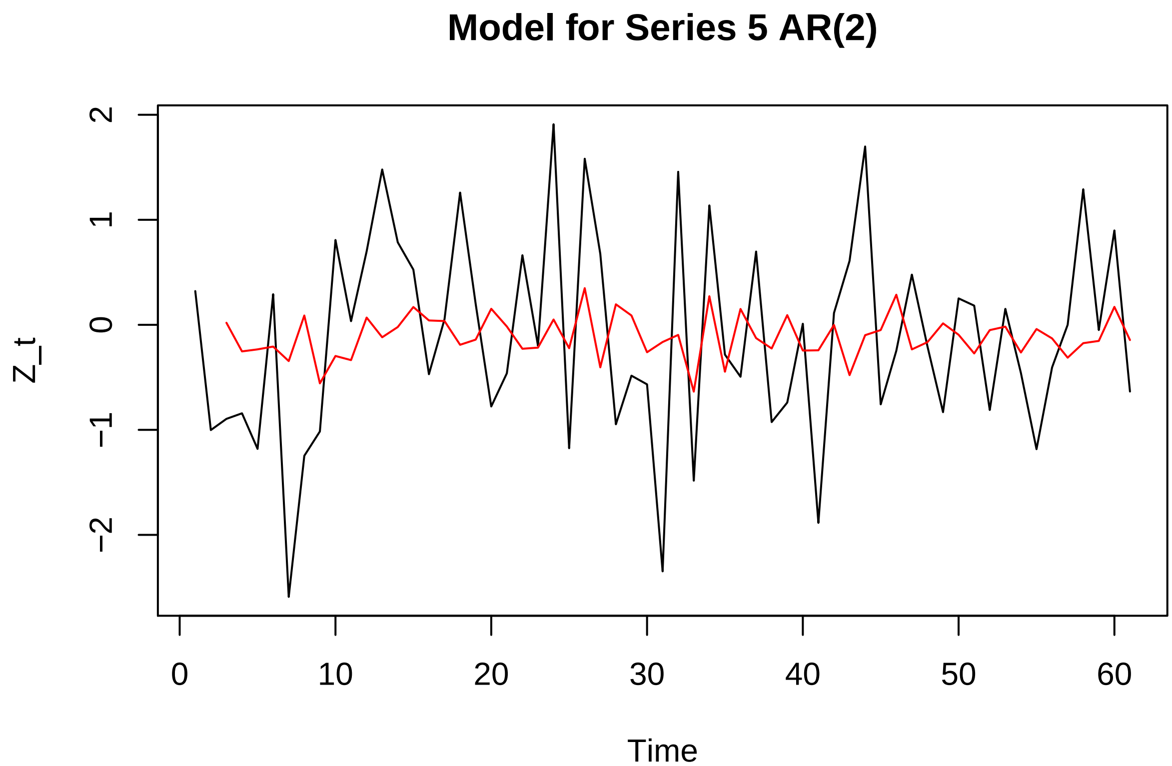






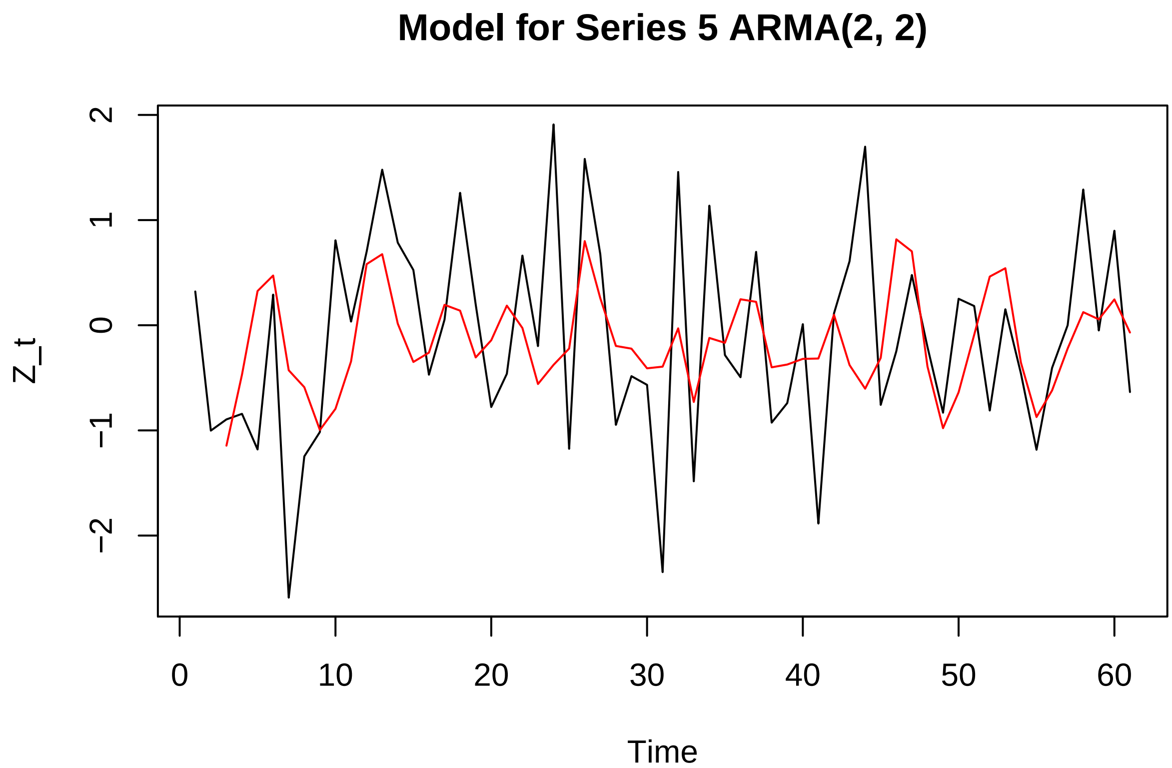
Vemos que tanto a ACF quanto a PACF são extremamente baixas para valores diferentes de 0, no entanto, nessa série possuímos bem menos amostras que as demais, possuindo apenas 61 amostras. Por esse motivo, iremos considerar que a importância do segundo lag na PACF e avaliar dois modelos distintos, AR(2) e ARMA(2, 2). Considerando primeiro o modelo AR(2).

```
##
## Call:
## arma(x = X[[5]], order = c(2, 0))
##
## Model:
## ARMA(2,0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24498 -0.70396  0.01893  0.71961  1.85873
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ar1        -0.04726     0.12481   -0.379    0.705
## ar2         0.20225     0.12576    1.608    0.108
## intercept  -0.09259     0.12355   -0.749    0.454
##
## Fit:
## sigma^2 estimated as 0.9154,  Conditional Sum-of-Squares = 53.09,  AIC = 173.72
```

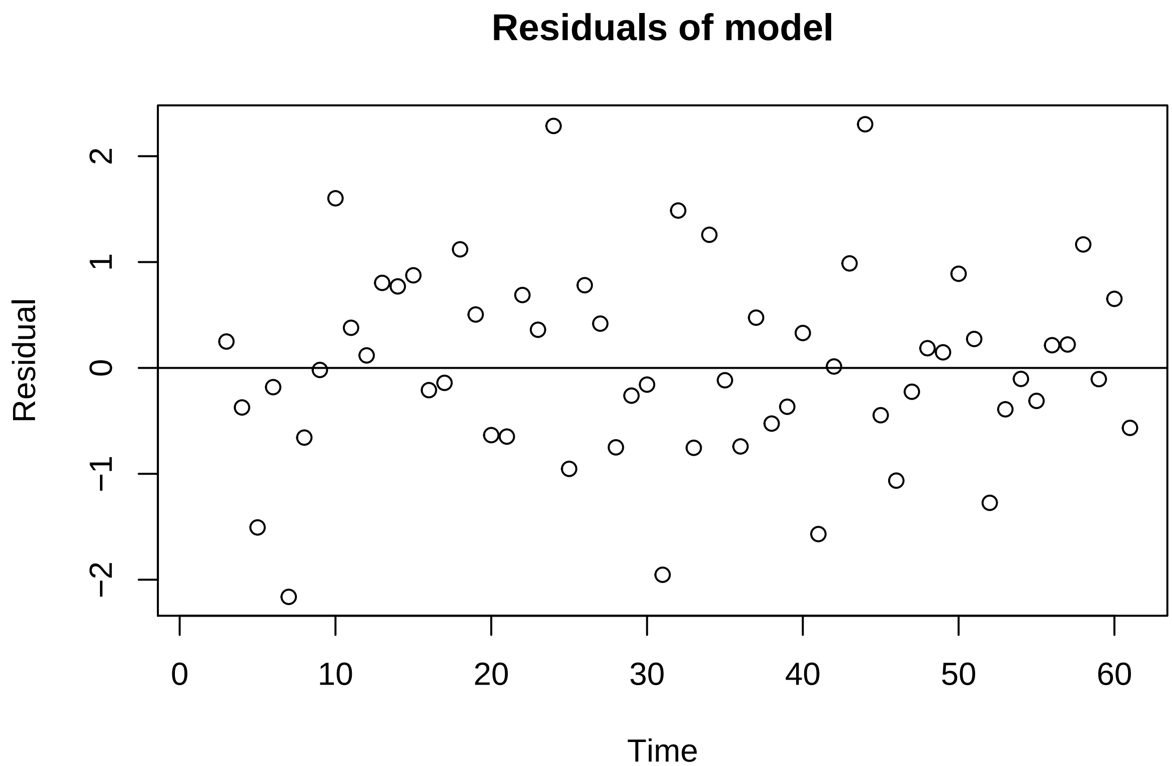


E agora o modelo ARMA(2, 2).

```
##
## Call:
## arma(x = X[[5]], order = c(2, 2))
##
## Model:
## ARMA(2,2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -2.16208 -0.48610 -0.01902  0.57863  2.30147
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ar1         0.81736     0.17739    4.608 4.07e-06 ***
## ar2        -0.62001     0.10722   -5.783 7.35e-09 ***
## ma1        -0.92912     0.09675   -9.603  < 2e-16 ***
## ma2         0.95619     0.10319    9.266  < 2e-16 ***
## intercept  -0.12685     0.11679   -1.086    0.277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 0.8179,  Conditional Sum-of-Squares = 47.54,  AIC = 170.85
```

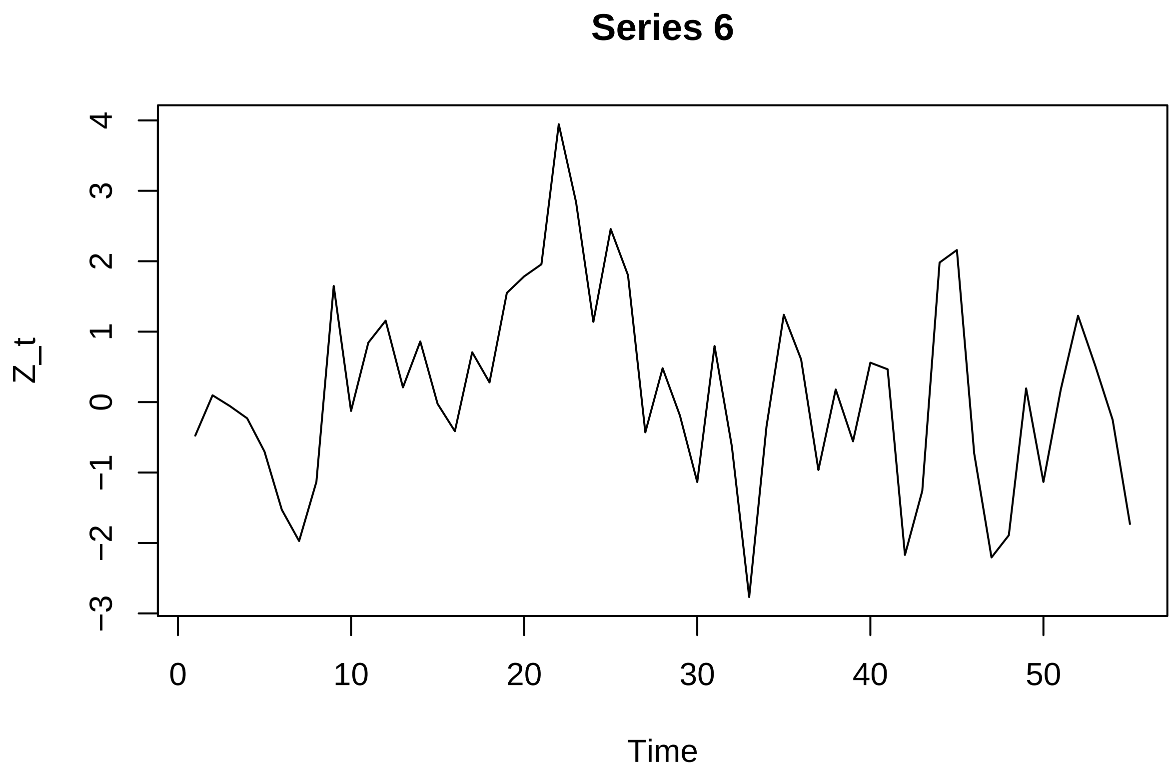


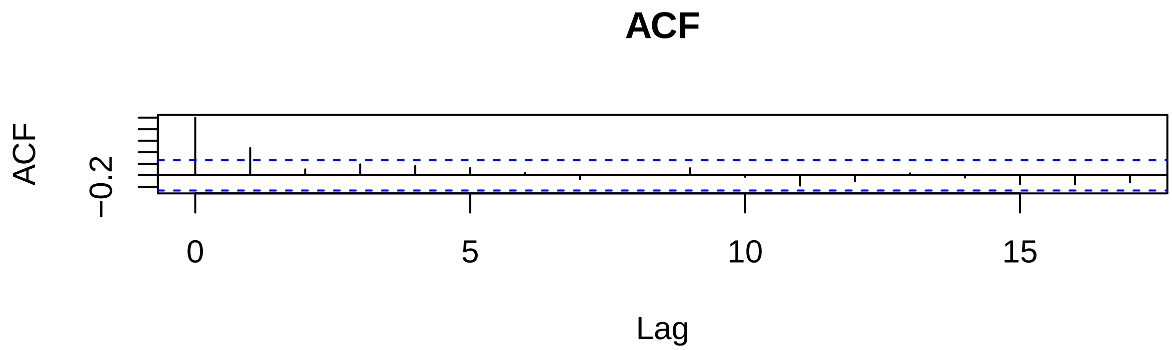
Vemos que o modelo ARMA(2, 2) apresentou um AIC menor, de 170, em comparação com o AR(2), além disso, ele também encaixou melhor na curva real dos dados. Vamos visualizar os resíduos:

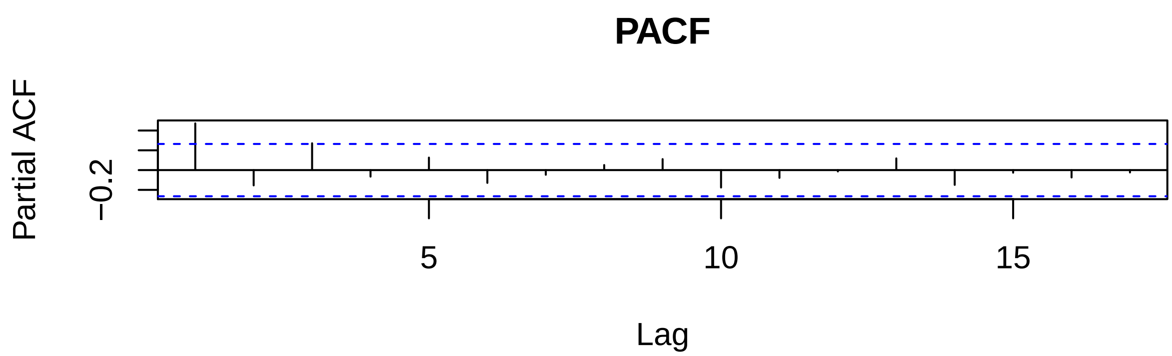


Residuals of model
Residual
2
1
0
−1
−2
0
10
20
30
40
50
60
Time

Série 6

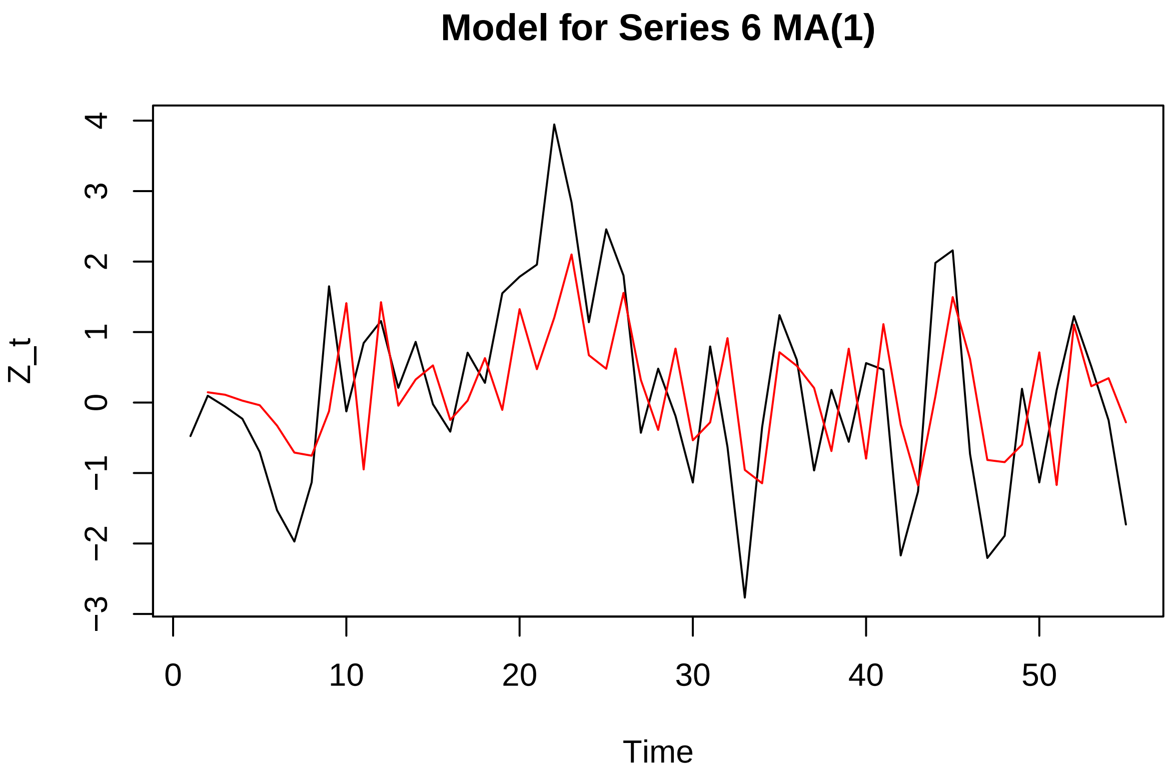


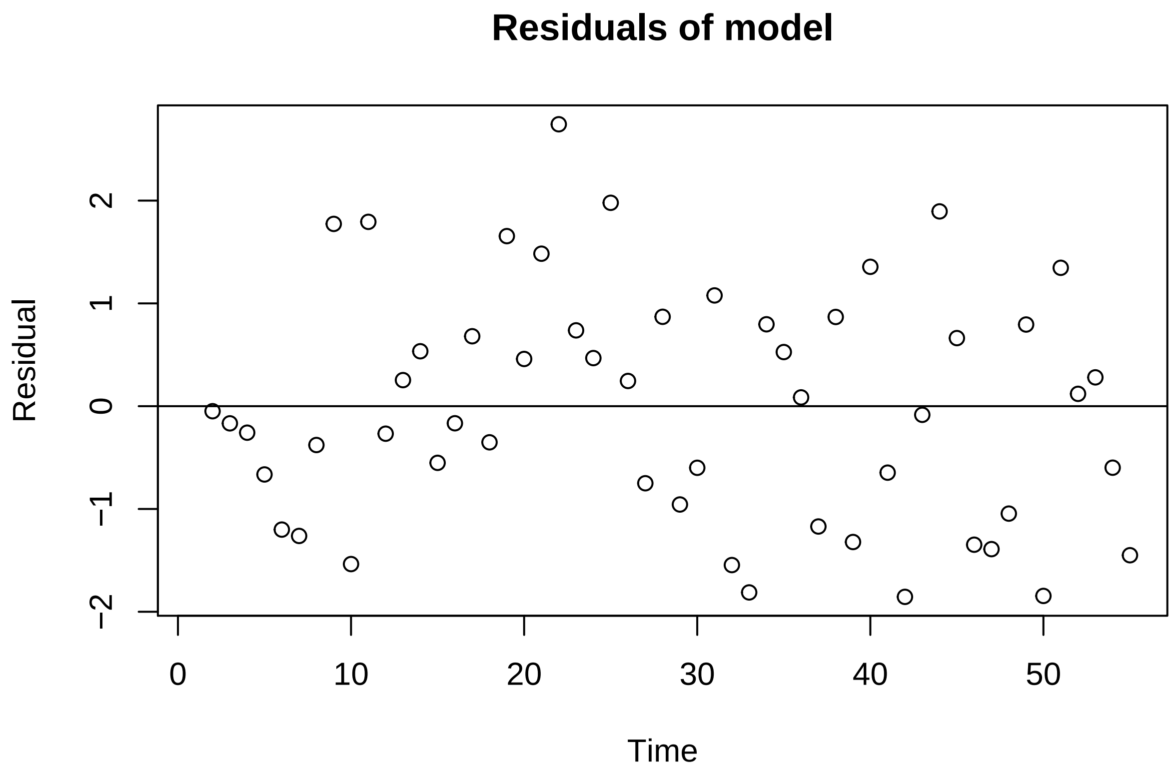


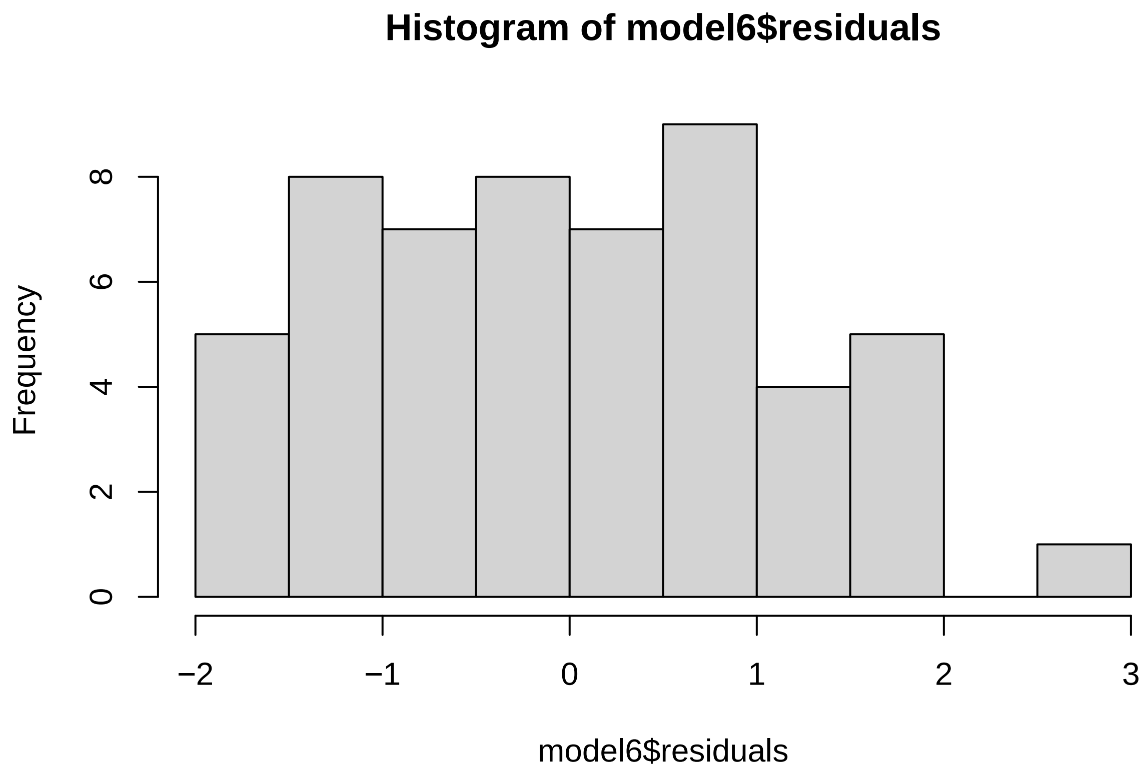


Aqui temos que o ACF diminui bastante para lag acima de 1 enquanto que o PACF se mantém consistentemente baixo para lag maiot que 0, o que indica que o modelo MA(1) pode ser apropriada para essa série.

```
##
## Call:
## arma(x = X[[6]], order = c(0, 1))
##
## Model:
## ARMA(0,1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85487 -0.90462 -0.06636  0.78021  2.74265
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ma1          0.7129      0.0991    7.194 6.31e-13 ***
## intercept    0.1458      0.2557    0.570    0.569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 1.29,  Conditional Sum-of-Squares = 68.38,  AIC = 174.1
```

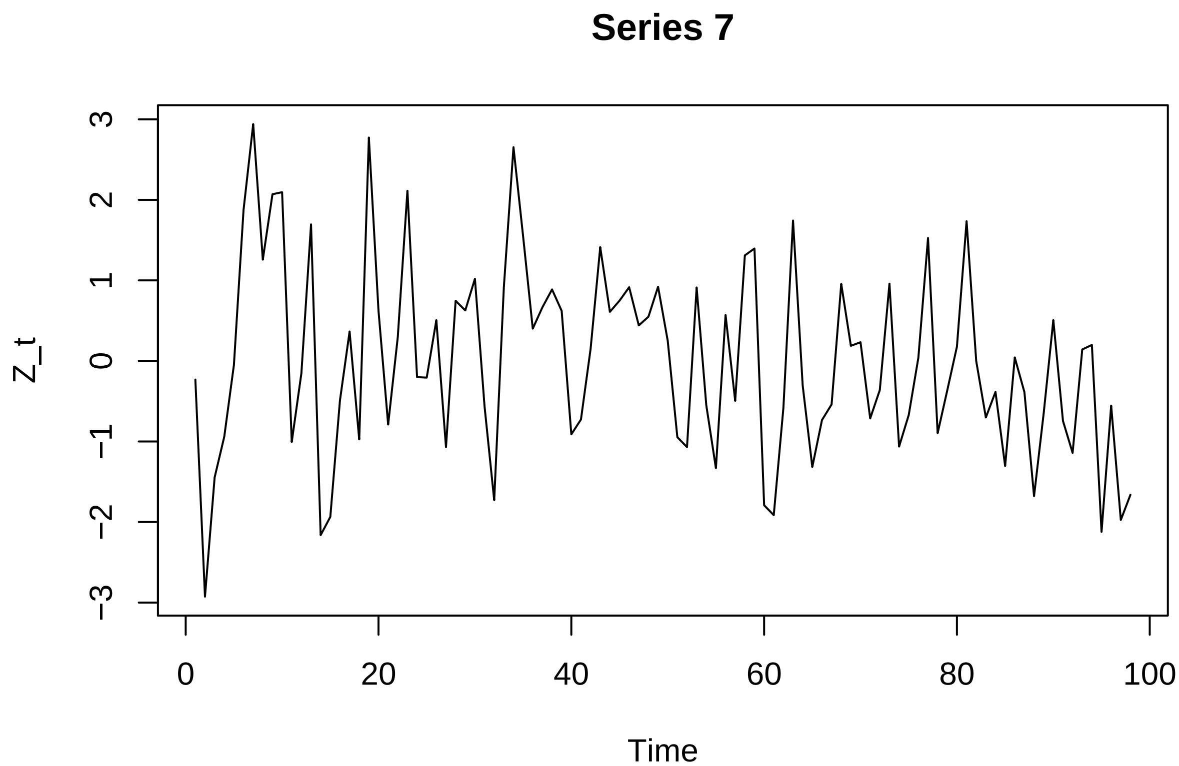

Model for Series 6 MA(1)
Z_t
Time
4
3
2
1
0
−1
−2
−3
0
10
20
30
40
50

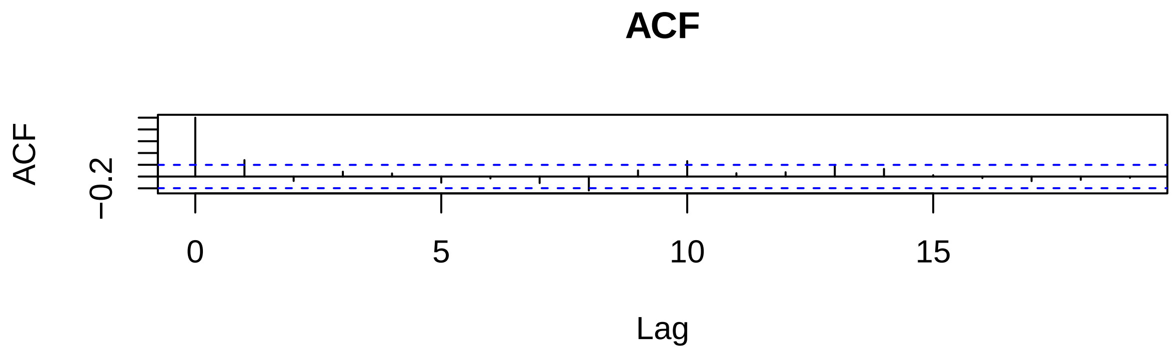

Residuals of model
Residual
−2
−1
0
1
2
0
10
20
30
40
50
Time

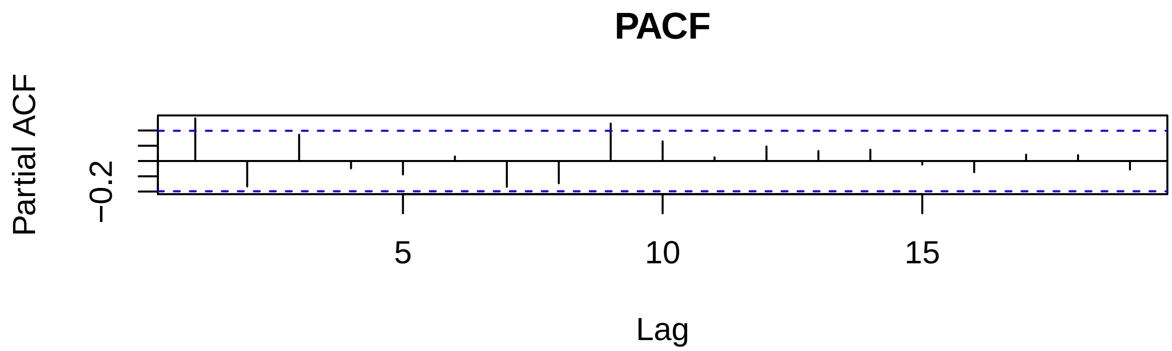


O modelo não conseguiu capturar bem os picos da série real, mas se encaixou razoavelmente bem.

Série 7

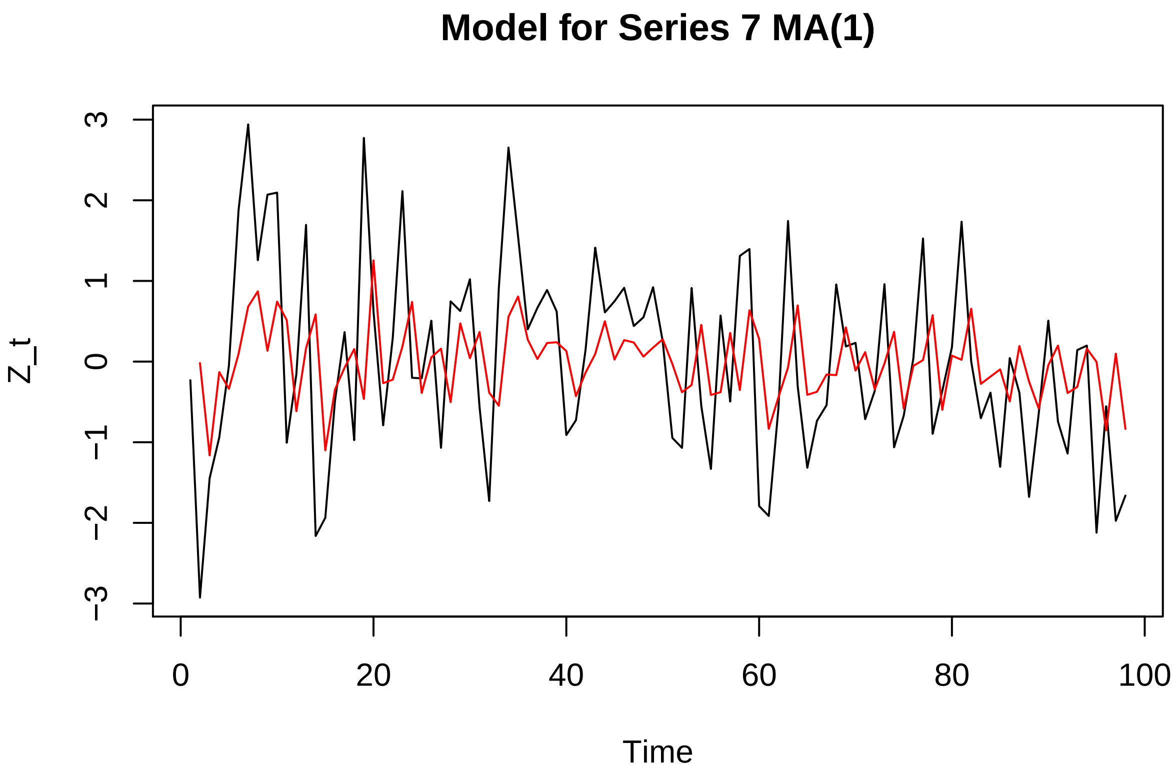


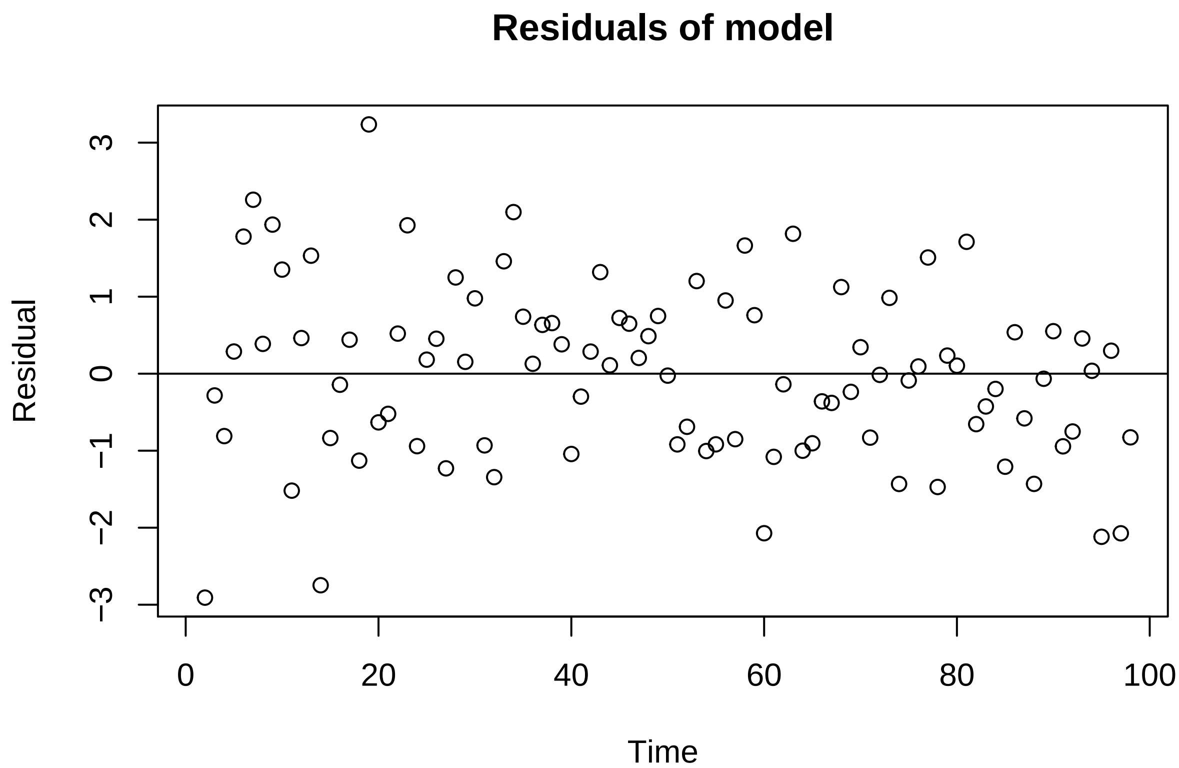


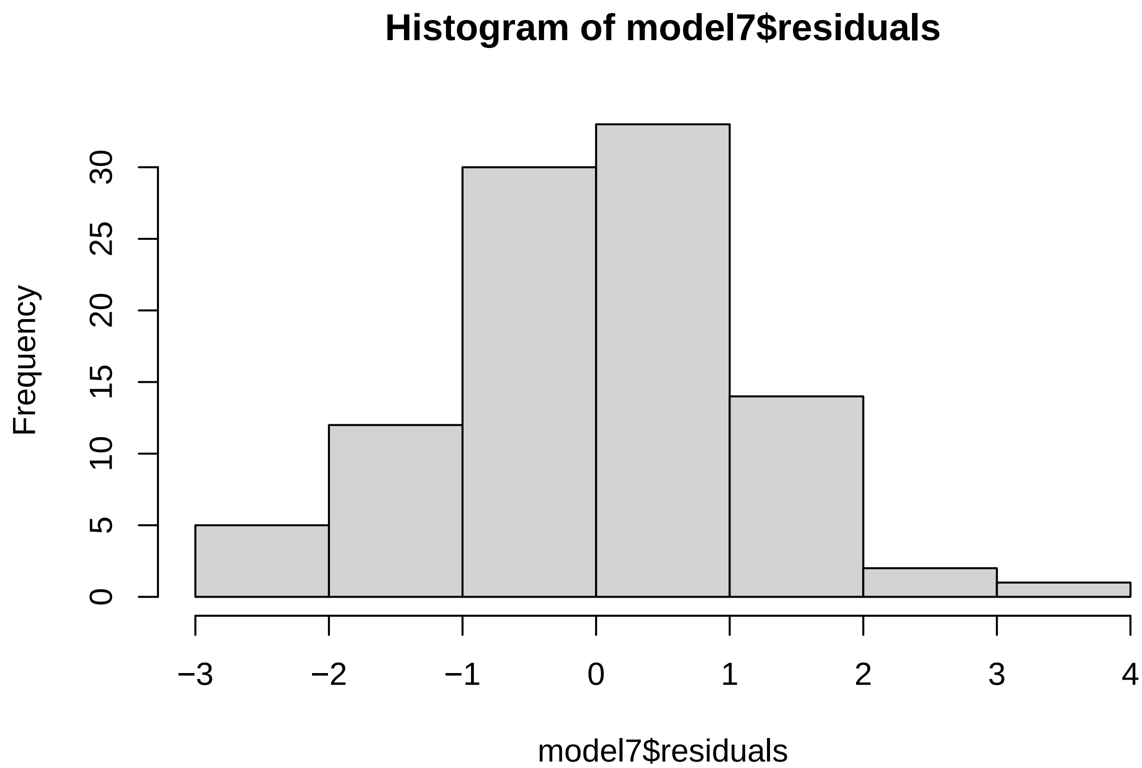


O ACF decai rapidamente para lag maior do que 1, enquanto que o PACF decresce bem lentamente, o que indica que o modelo MA(1) é uma boa escolha aqui.

```
##
## Call:
## arma(x = X[[7]], order = c(0, 1))
##
## Model:
## ARMA(0,1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.90811 -0.83574  0.09345  0.65703  3.23671
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ma1         0.39365     0.10095    3.900 9.64e-05 ***
## intercept  -0.01875     0.15811   -0.119    0.906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 1.282,  Conditional Sum-of-Squares = 123.1,  AIC = 306.47
```

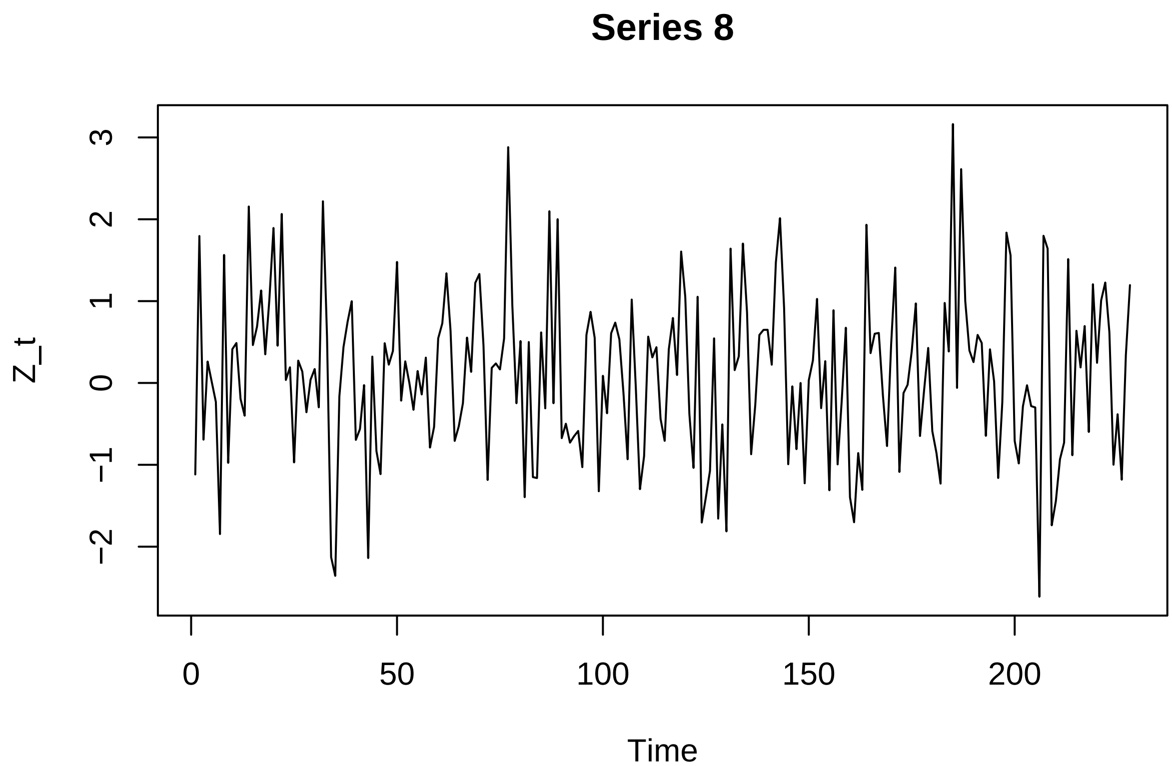

Model for Series 7 MA(1)
Z_t
3
2
1
0
−1
−2
−3
0
20
40
60
80
100
Time

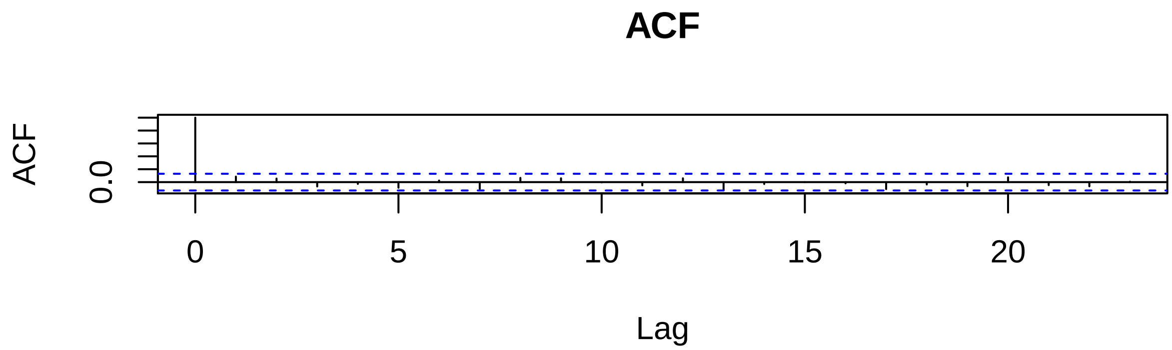

Residuals of model
Residual
3
2
1
0
−1
−2
−3
0
20
40
60
80
100
Time

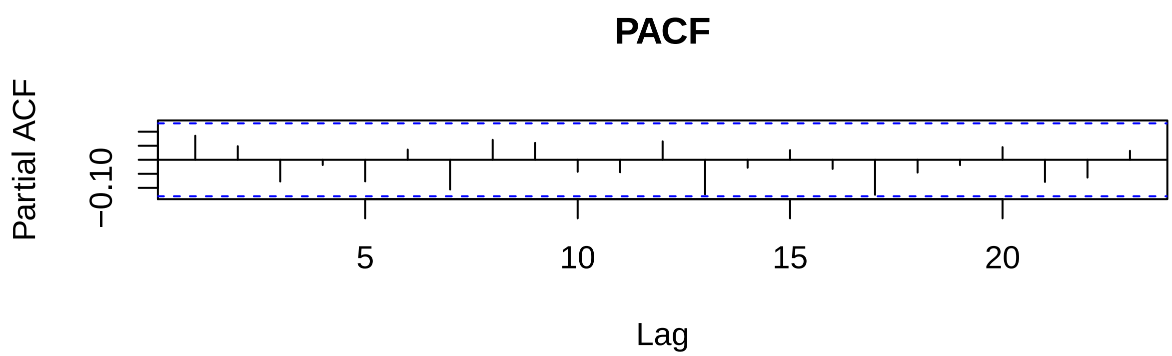


O modelo se enccaixou razoavelmente aos dados, mas não capturou bem os picos apresentados.

Série 8



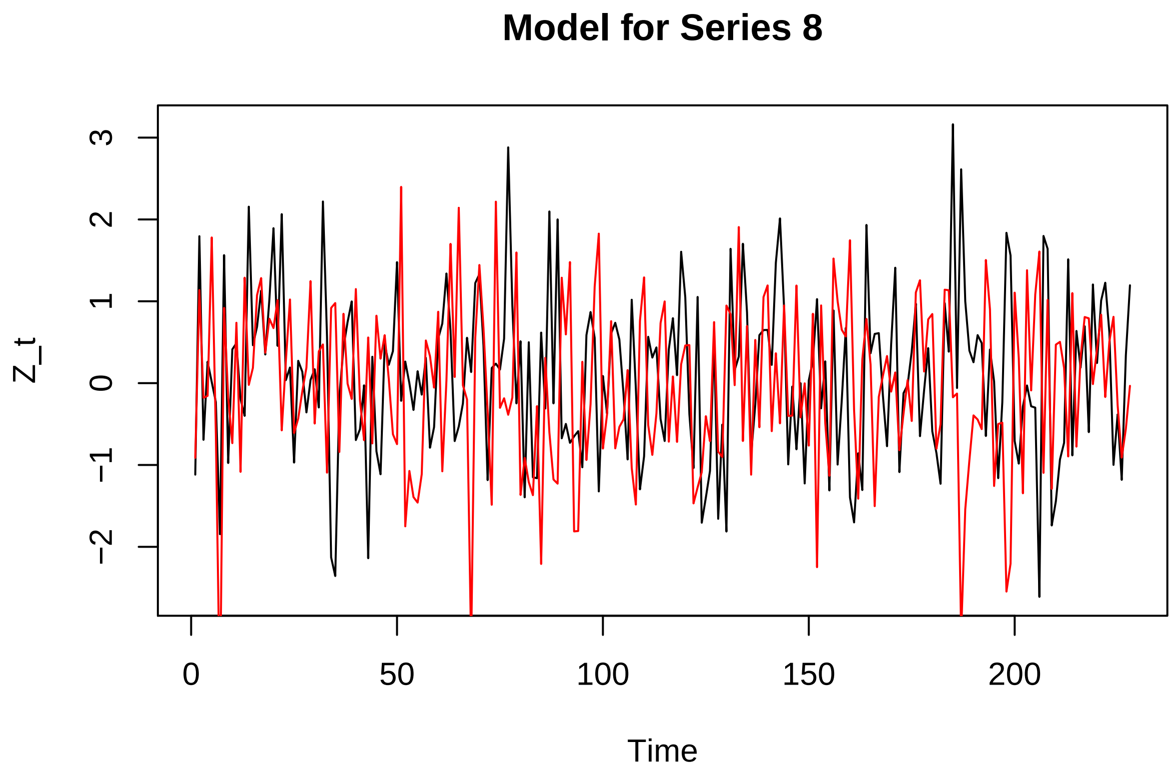




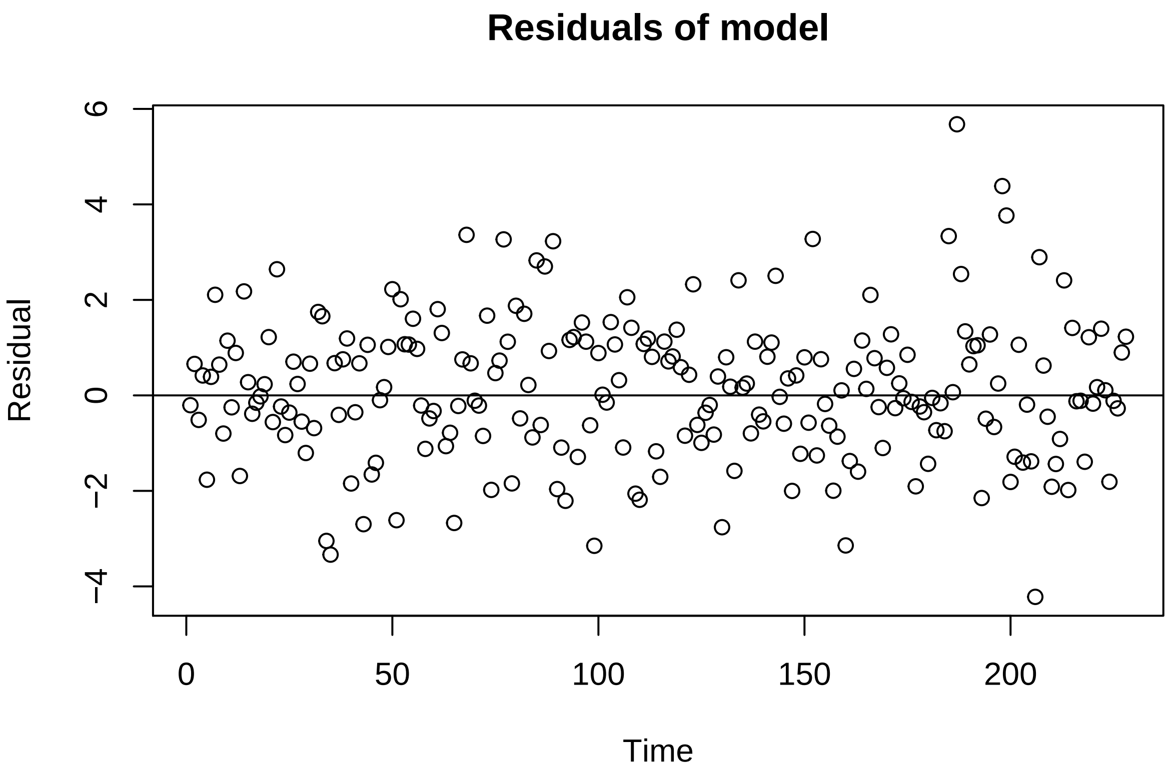
Tanto ACF quanto PACF são bem pequenas para lag maiores que zero, indicando que é um ruído branco.

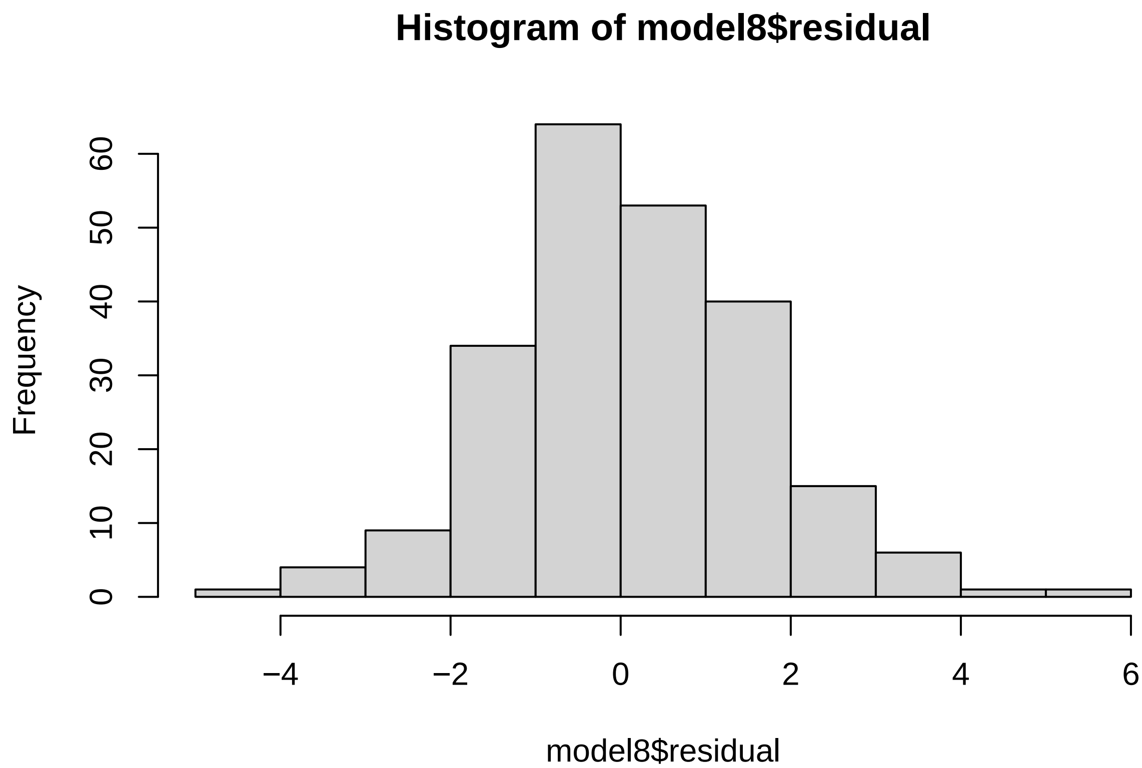
```
##         mean variance
## 1 0.09216291 0.996732
```

Temos que $E(a_t) = 0$ e $Var(a_t) = 1$, o que indica que o modelo se comporta como um ruído branco. Gerarmos 228 amostras de a_t e visualizarmos tanto o modelo e predição, quanto o residual:

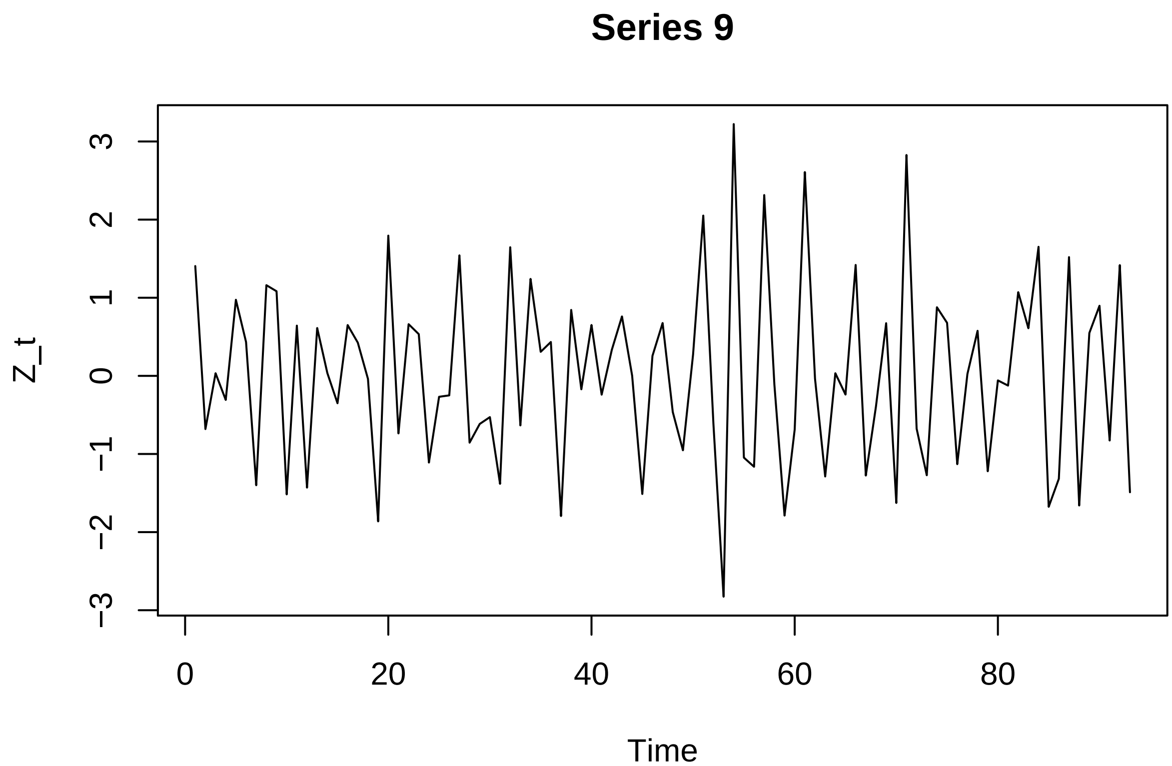


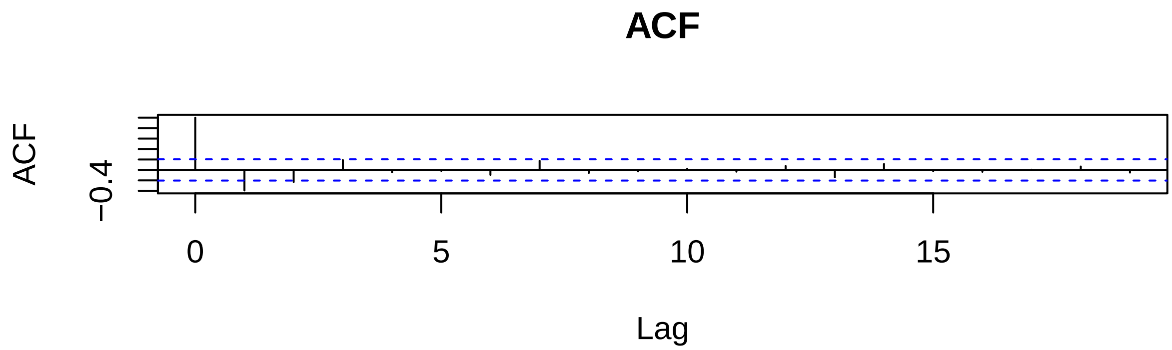

```
## Warning in model8$residual <- X[[8]] - model8: Realizando coerção de LHD para
## uma lista
```

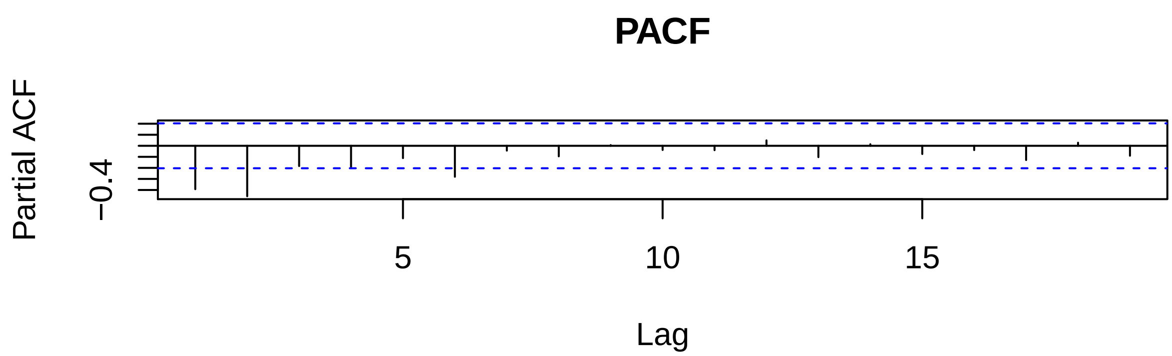

Residuals of model
Residual
Time
6
4
2
0
−2
−4
0
50
100
150
200


Histogram of model8$residual
Frequency
0
10
20
30
40
50
60
−4
−2
0
2
4
6
model8$residual

Série 9

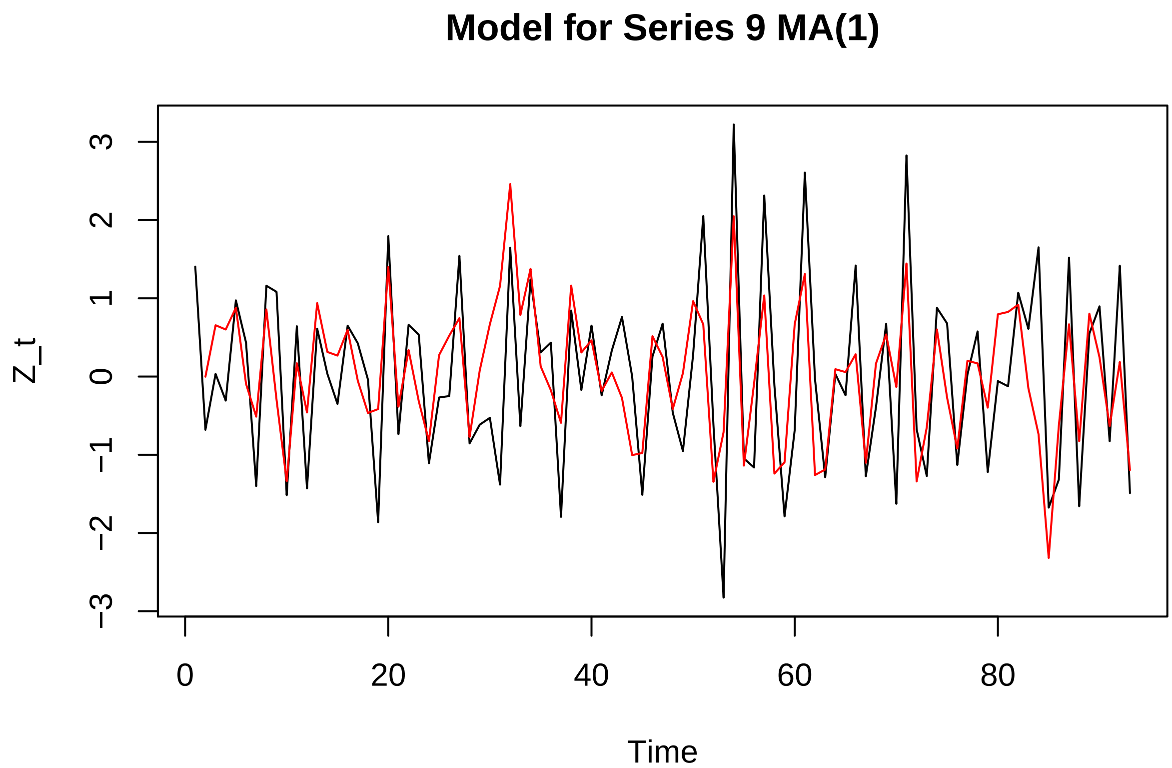




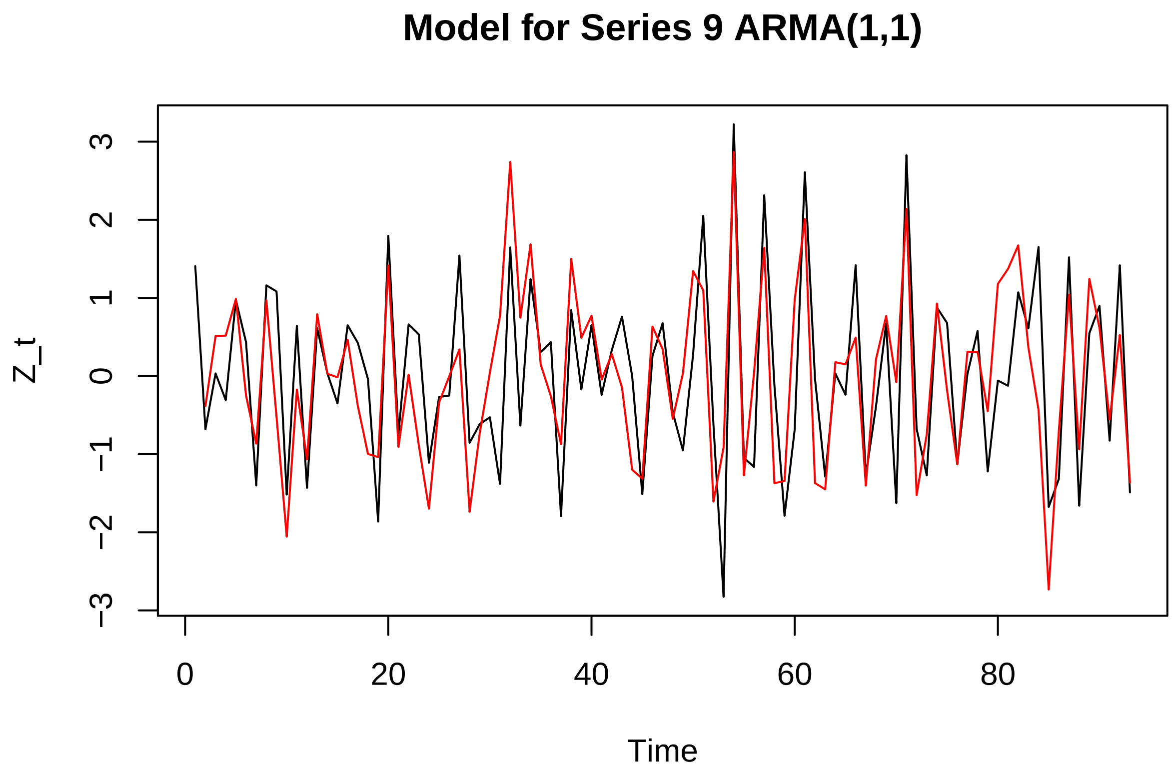


O ACF decai rapidamente para lag maior do que 1, equanto que o PACF decai mais lentamente para lag maior do que 1. Assim, testaremos os modelos MA(1) e ARMA(1,1).

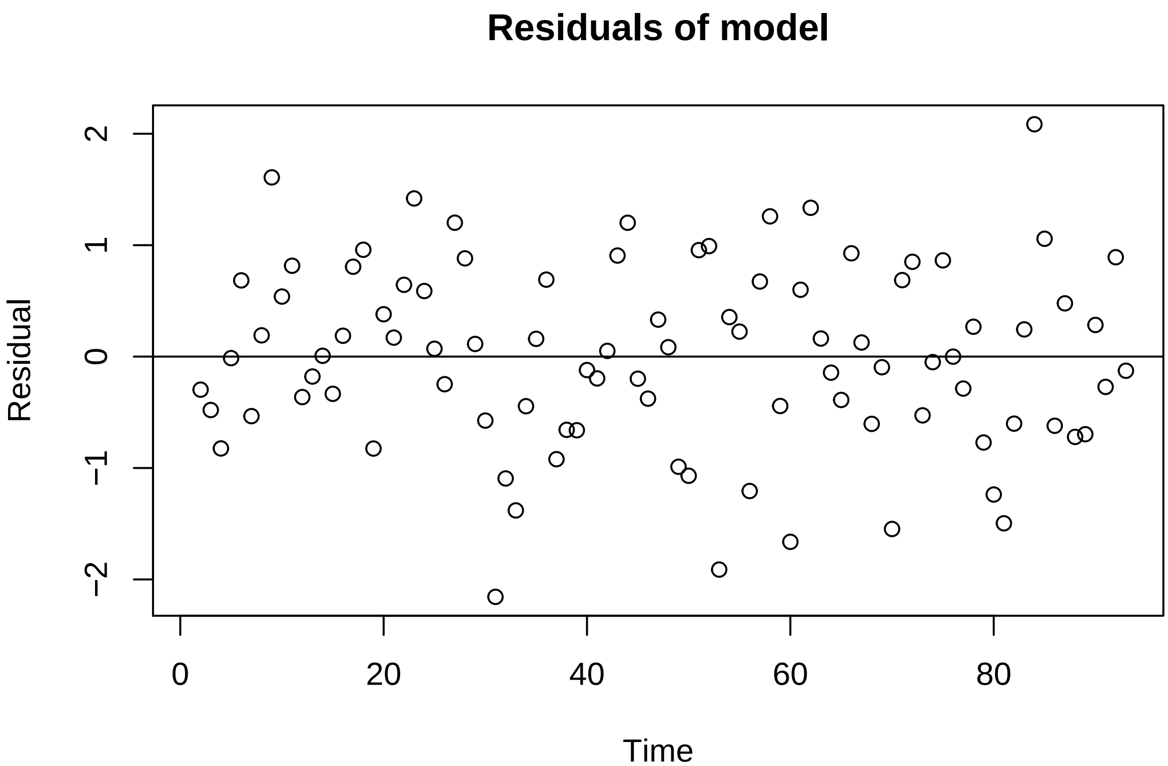
```
##
## Call:
## arma(x = X[[9]], order = c(0, 1))
##
## Model:
## ARMA(0,1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5405 -0.6815 -0.1173  0.5474  2.3902
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ma1       -0.969369    0.038766  -25.006   <2e-16 ***
## intercept -0.002035    0.004299   -0.473    0.636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 0.7455,  Conditional Sum-of-Squares = 68.05,  AIC = 240.61
```




```
## Warning in arma(X[[9]], order = c(1, 1)): Hessian negative-semidefinite

## Warning in sqrt(diag(object$vcov)): NaNs produzidos

## Warning in sqrt(diag(object$vcov)): NaNs produzidos

##
## Call:
## arma(x = X[[9]], order = c(1, 1))
##
## Model:
## ARMA(1,1)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -2.156261 -0.544383  0.003433  0.651768  2.085140
##
## Coefficient(s):
##            Estimate  Std. Error  t value Pr(>|t|)
## ar1       -0.274298          NA       NA       NA
## ma1       -1.094195          NA       NA       NA
## intercept  0.001394          NA       NA       NA
##
## Fit:
## sigma^2 estimated as 0.6787,  Conditional Sum-of-Squares = 61.76,  AIC = 233.87
```



Tivemos que o modelo ARMA(1,1) apresentou AIC menor que o modelo MA(1), além de aparentar visualmente se encaixar um pouco melhor nos dados.


Residuals of model
Residual
Time
2
1
0
−1
−2
0
20
40
60
80

O modelo se encaixou bem aos dados, conseguindo capturar razoavelmente os picos.