

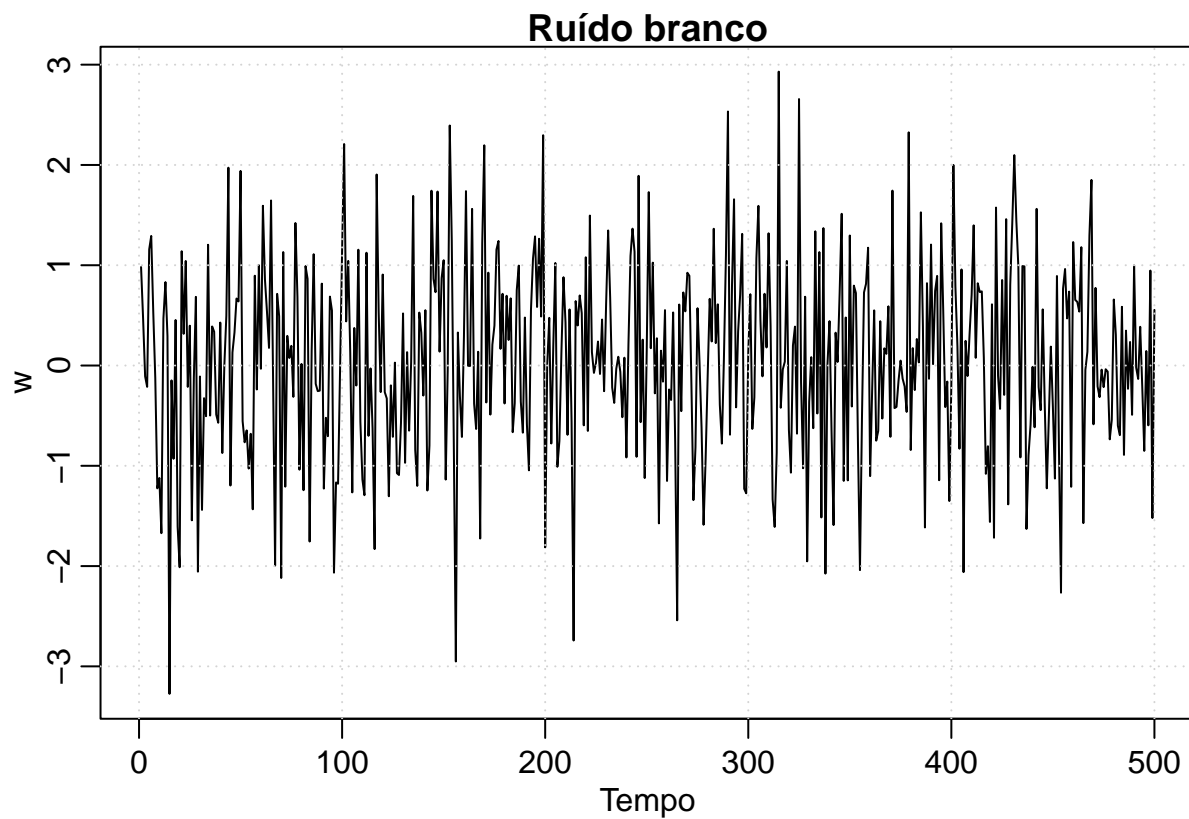
CE017 - Análise de Séries Temporais

Trabalho 1 - Luiz Henrique Barretta Francisco - GRR20213026

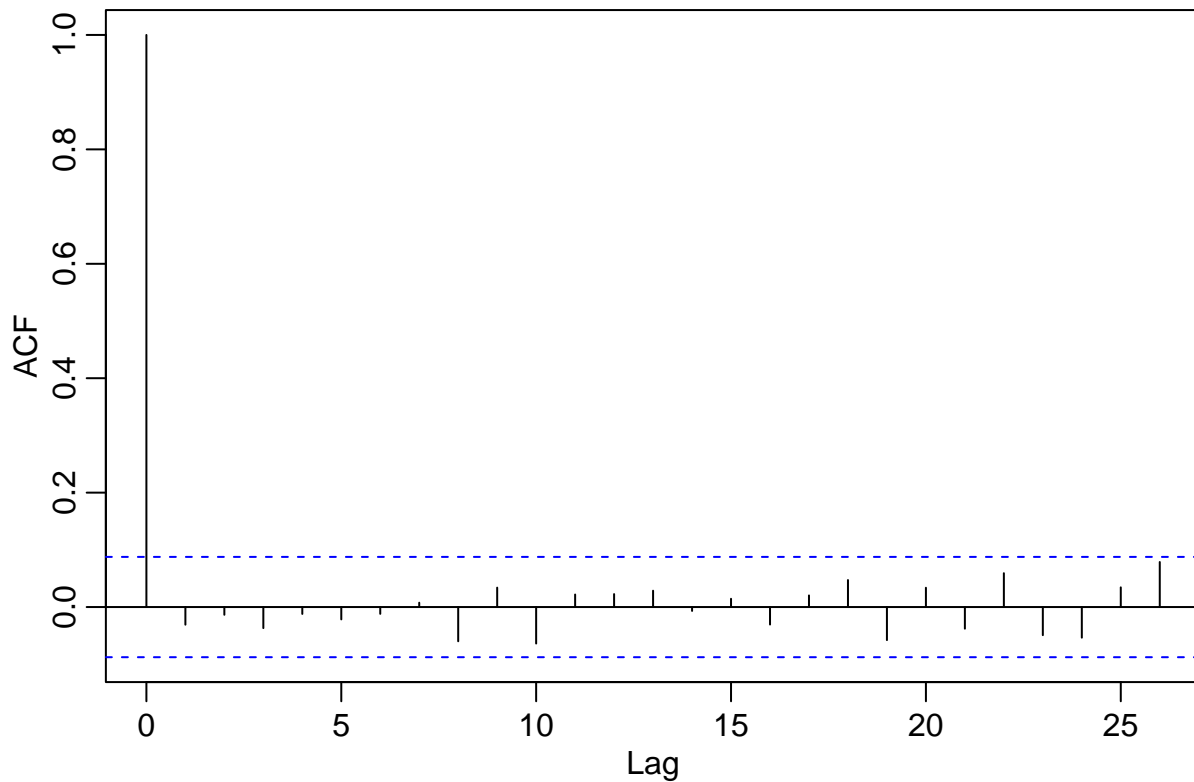
```
install.packages("astsa")  
library(astsa)
```

17- a) Simule uma série de $n=500$ observações de um processo ruído branco Gaussiano como no Exemplo 8 e calcule a função ACF mostral $\hat{\rho}(h)$, para até o lag 20. Compare a ACF amostral que você obteve com o ACF real $\rho(h)$.

```
set.seed(2024)  
w = rnorm(500,0,1) # 500  $N(0,1)$  variáveis  
par(mfrow=c(1,1),mar=c(4,3,1,1),mgp=c(1.6,.6,0))  
plot.ts(w, xlab="Tempo", main="Ruído branco")  
grid()
```



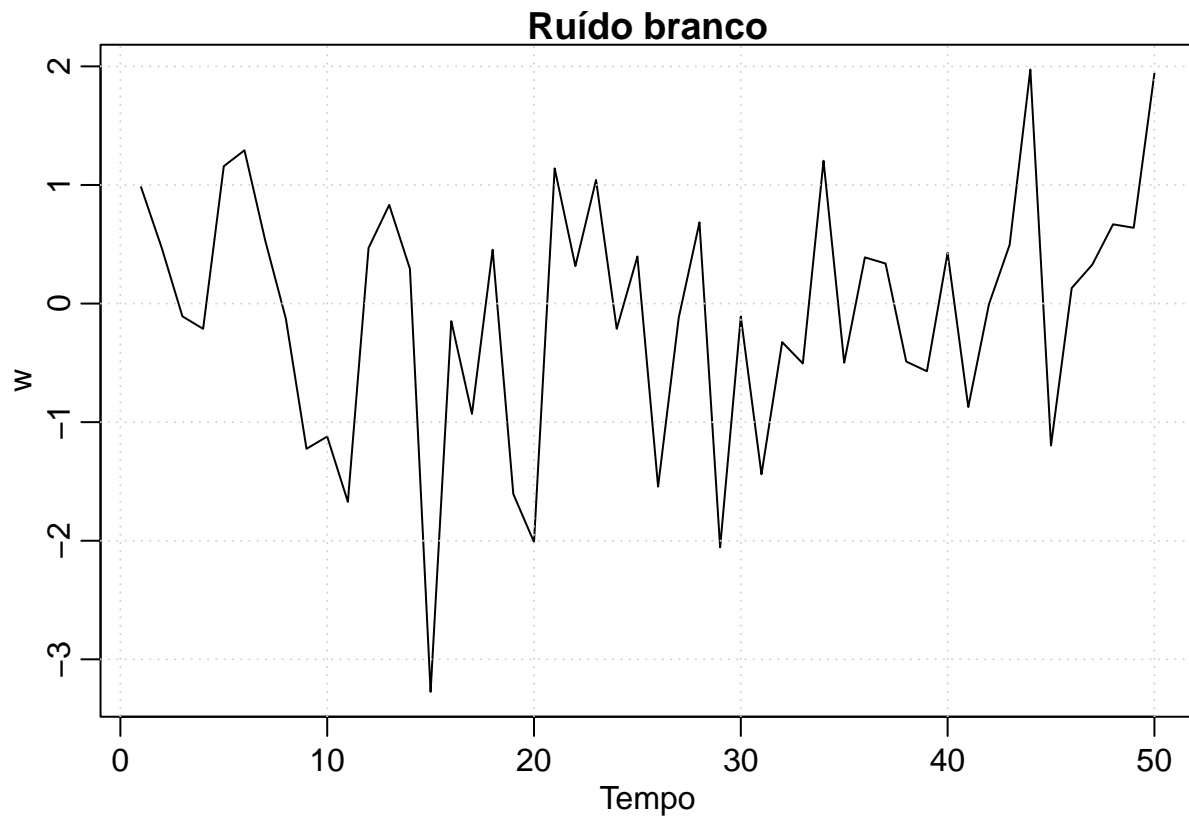
```
acf(w)
```



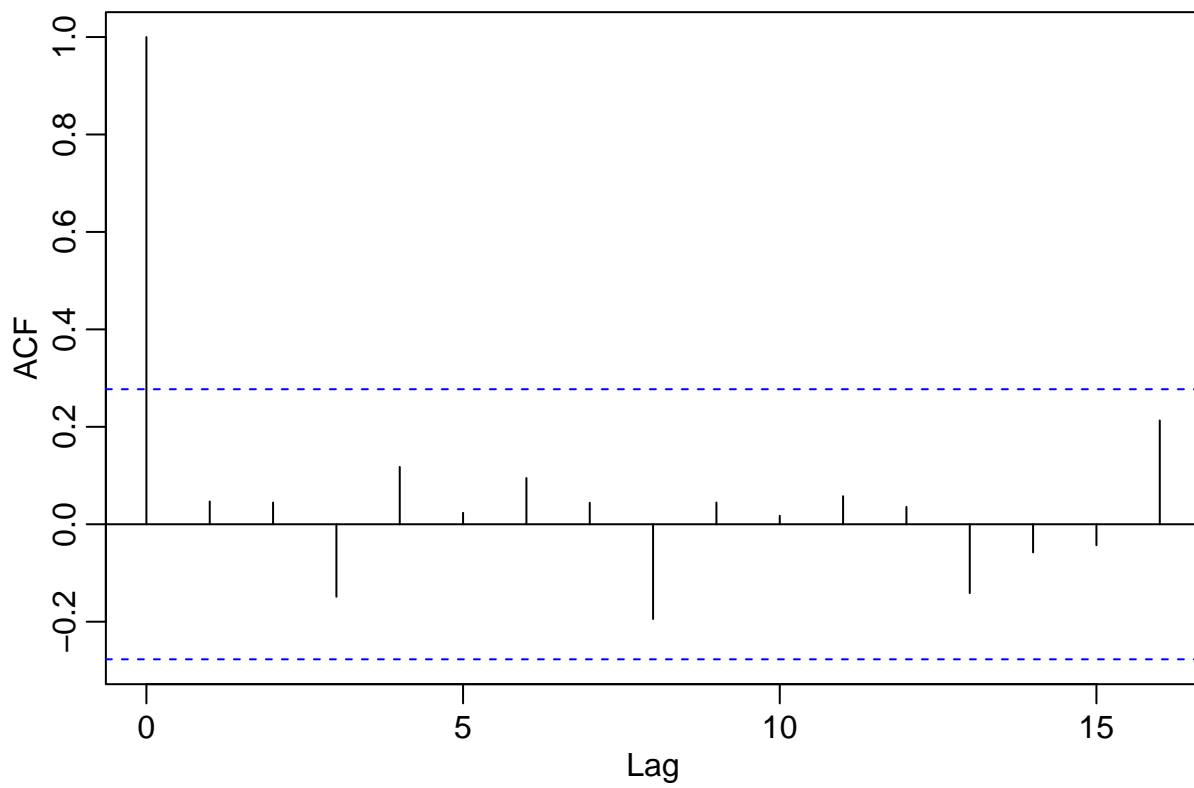
A ACF real de um ruído branco Gaussiano é $\rho(h) = 1$ se $h = 0$ e $\rho(h) = 0$ para qualquer $h \geq 1$. Esse foi exatamente o resultado obtido na ACF amostral mostrada no gráfico acima, dado que a autocorrelação da própria observação, quando $h = 0$, é $\hat{\rho}(h) = 1$ e $\hat{\rho}(h) = 0$ para todos os outros valores, pois estes ficaram dentro das bandas de confiança, e portanto devem ser considerados como 0.

b) Repita a parte (a) usando apenas $n=50$. Como mudar n afeta os resultados?

```
set.seed(2024)
w = rnorm(50,0,1) # 50 N(0,1) variáveis
par(mfrow=c(1,1),mar=c(4,3,1,1),mgp=c(1.6,.6,0))
plot.ts(w, xlab="Tempo", main="Ruído branco")
grid()
```



`acf(w)`

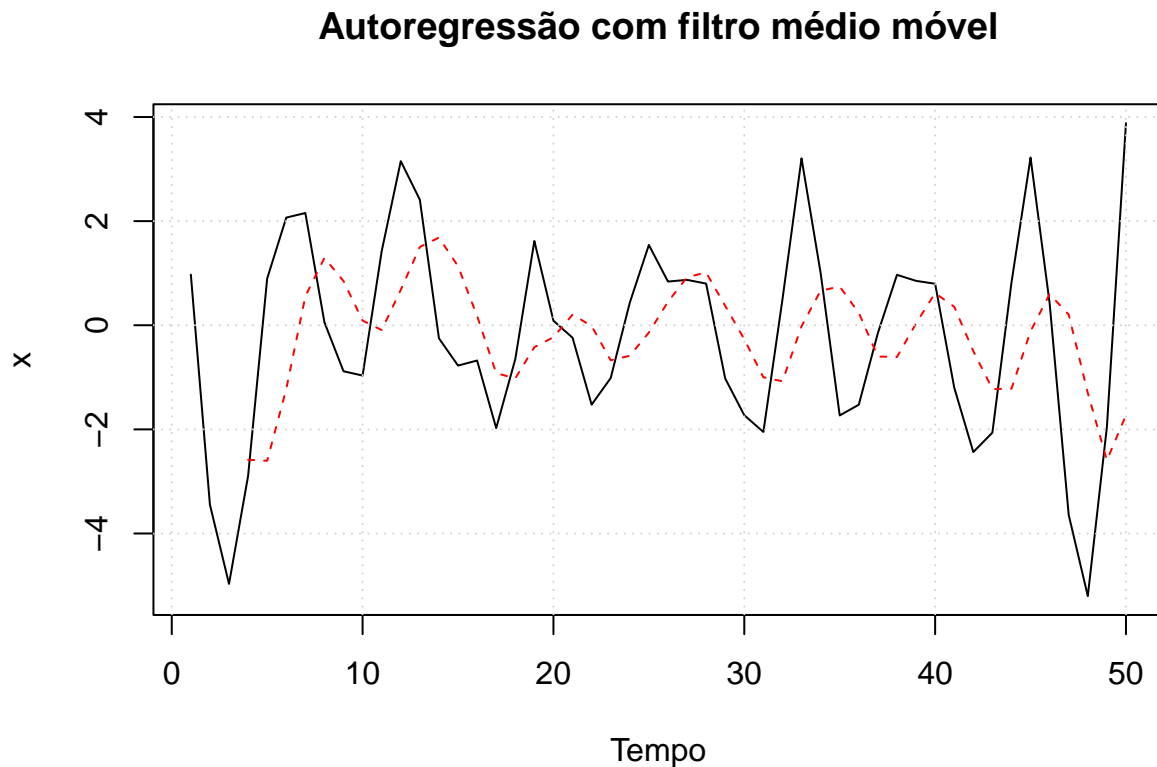


Com menos observações a ACF amostral tende a ser mais ruidosa e menos precisa quando comparada à

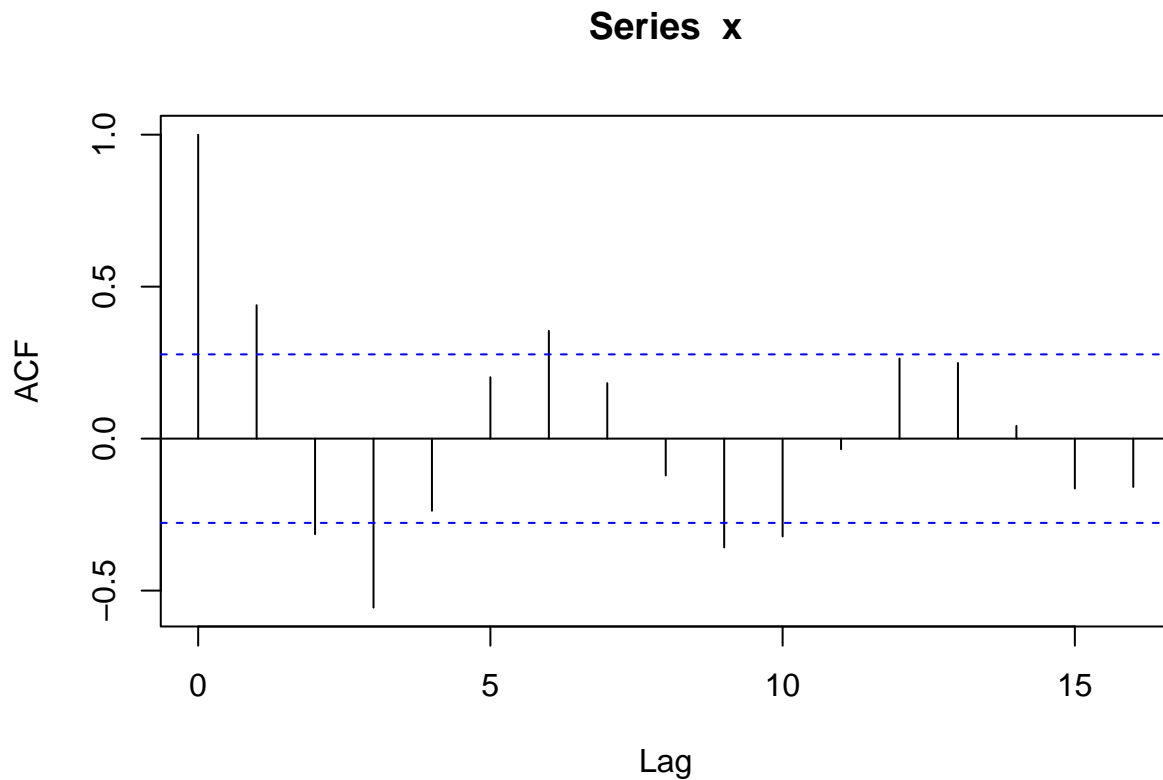
ACF real. Como mostrado no gráfico acima, que nitidamente tem bandas de confiança mais largas do que no exemplo com mais observações, nos dando menos confiança de que esses valores são zero, por mais que no exemplo da *seed* gerada ainda todos o sejam.

19- Embora o modelo no Exercício 2 (a) não seja estacionário, a função ACF amostral pode ser informativa. Para os dados gerados nesse problema, calcule e mostre a ACF amostral e, em seguida, comente.

```
set.seed(2024)
w = rnorm(100, 0, 1)
x = filter(w, filter = c(1, -0.9), method = "recursive")[-(1:50)]
v = filter(x, rep(1/4, 4), sides = 1)
plot.ts(x, xlab = "Tempo", main = "Autoregressão com filtro médio móvel")
lines(v, lty = 2, col = "red")
grid()
```



```
acf(x)
```

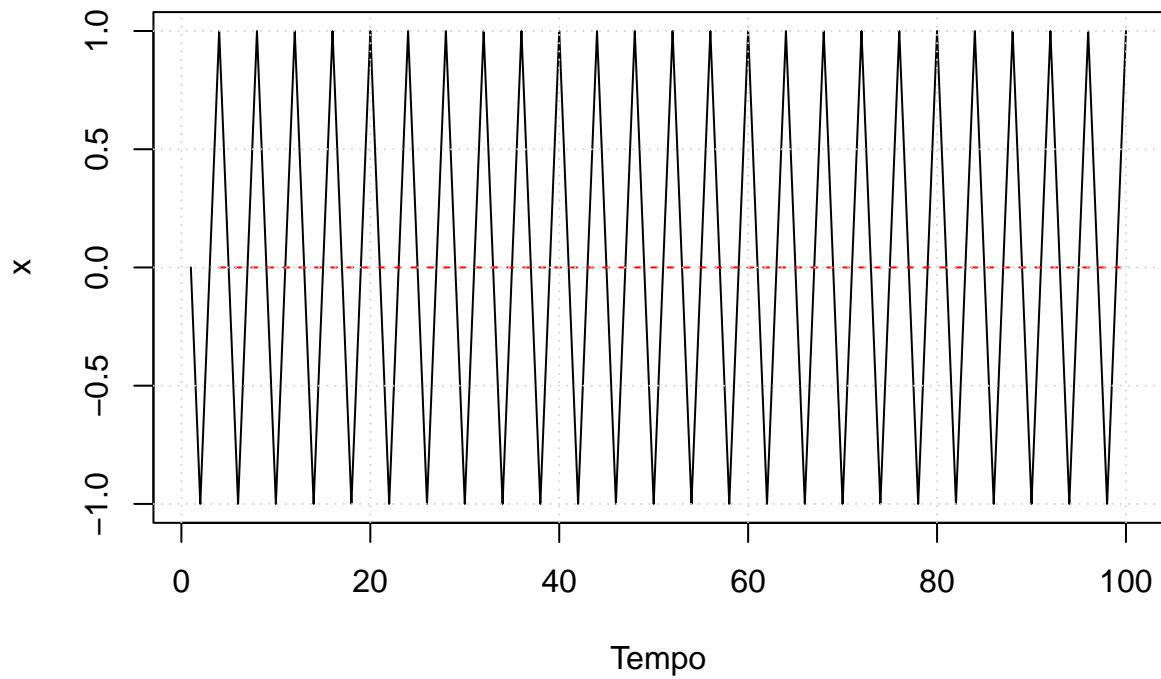


A ACF amostral mostra uma autocorrelação significativa nos primeiros valores e alternância a cada 3 ou 4 unidades de lag, o que corresponde com o modelo autorregressivo proposto.

```
set.seed(2024)
t = 1:100
x = cos(2 * pi * t / 4)
v = filter(x, rep(1/4, 4), sides = 1)

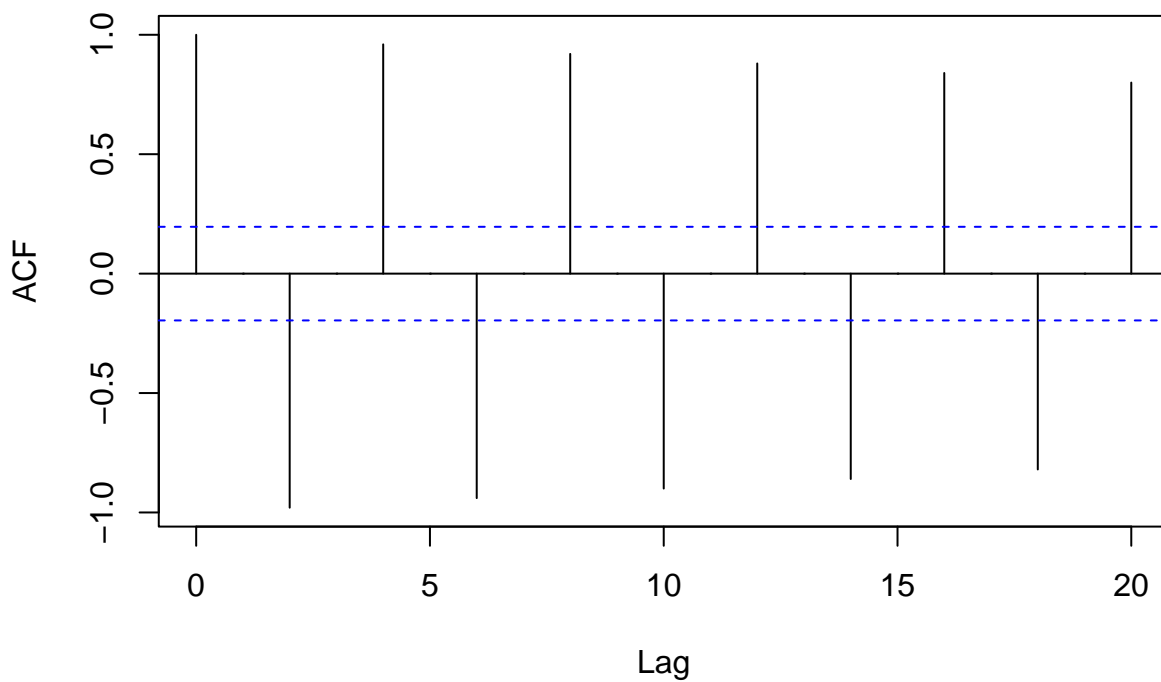
plot.ts(x, xlab = "Tempo", main = "Cosseno com filtro médio móvel")
lines(v, lty = 2, col = "red")
grid()
```

Cosseno com filtro médio móvel



`acf(x)`

Series x



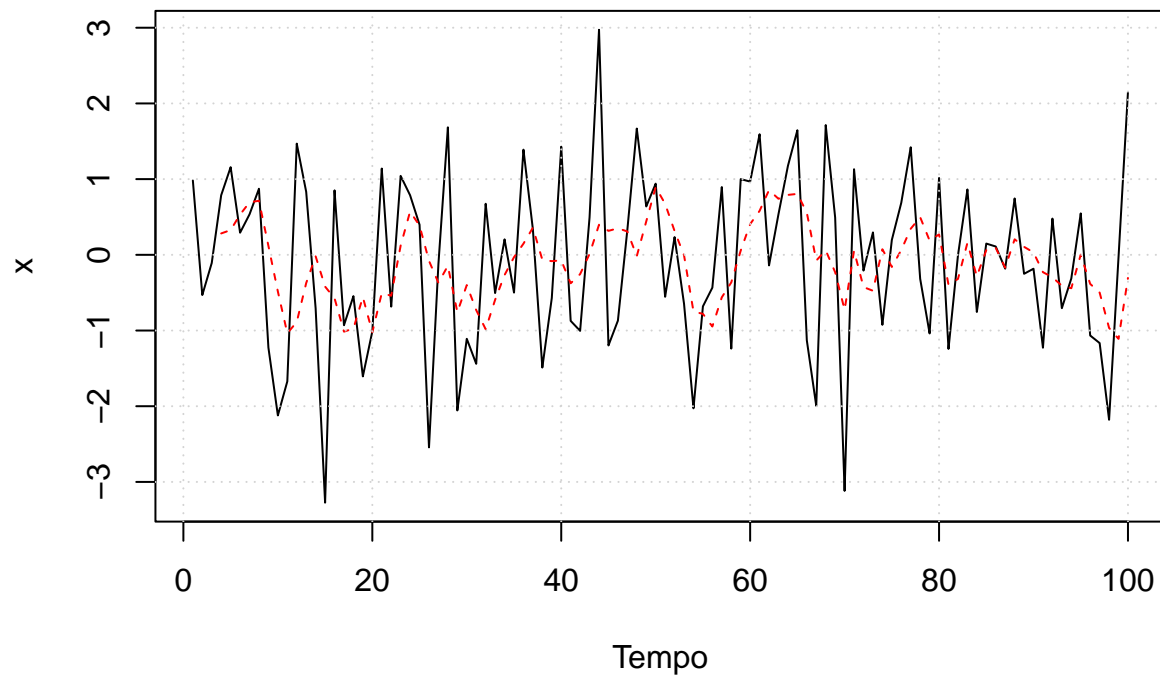
A ACF amostral apresenta um padrão cíclico e alternado a cada 4 valores, exatamente por ter sido construída da função periódica $\cos(2\pi t/4)$.

```

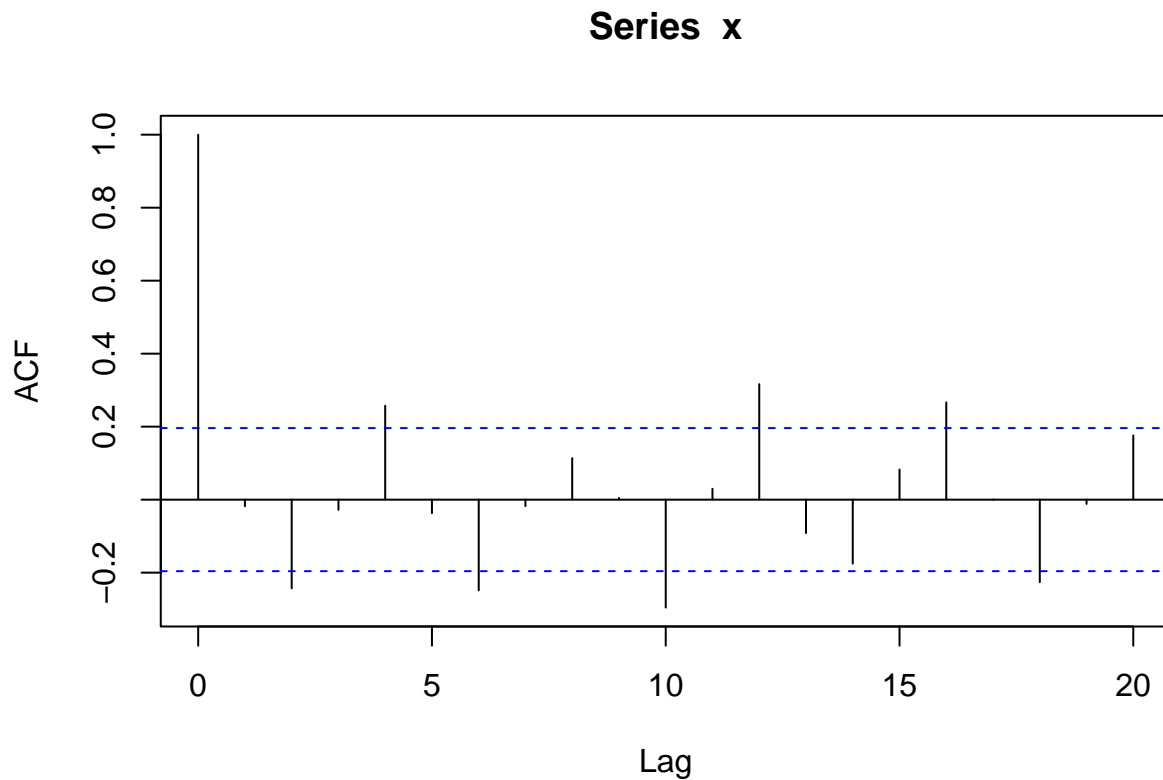
x = x + w
v = filter(x, rep(1/4, 4), sides = 1)
plot.ts(x, xlab = "Tempo", main = "Cosenoidal com ruído e filtro médio móvel")
lines(v, lty = 2, col = "red")
grid()

```

Cosenoidal com ruído e filtro médio móvel



```
acf(x)
```

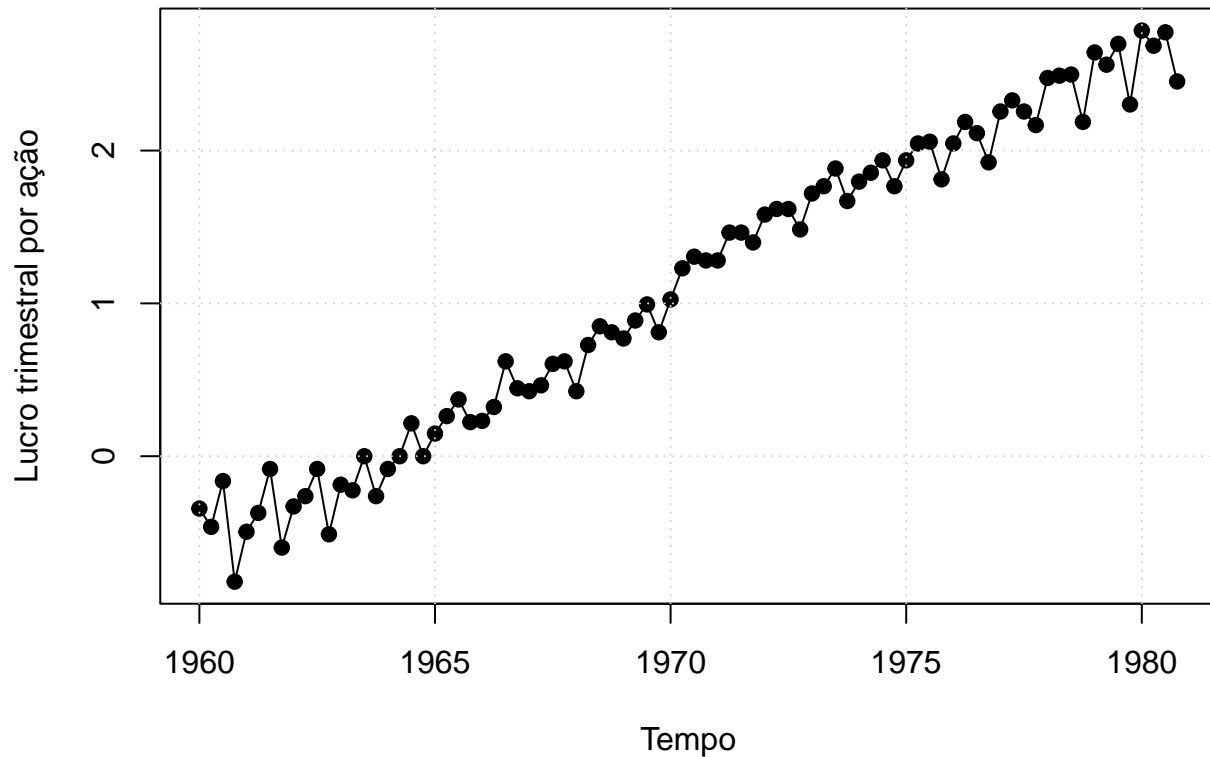


Ao adicionarmos um ruído branco Gaussiano na função periódica $\cos(2\pi t/4)$ a ACF amostral ainda apresenta seu padrão cíclico e alternado porém muito menos significativo dado a adição do ruído branco, já que ele tende a diminuir a autocorrelação.

1- Um Modelo Estrutural: Para os dados da Johnson & Johnson y_t , mostrados na figura considerando $x_t = \log(y_t)$.

```
par(mar=c(4,4,3,1))
plot(log(jj), type="o", xlab="Tempo", ylab="Lucro trimestral por ação", pch=19,
     main="Johnson & Johnson ganhos trimestrais por ação\n 84 trimestres, 1960-I a 1980-IV")
grid()
```


Johnson & Johnson ganhos trimestrais por ação 84 trimestres, 1960-I a 1980-IV



Neste problema, vamos ajustar um tipo especial de modelo estrutural, $X_t = T_t + S_t + W_t$, em que T_t é um componente de tendência, S_t um componente sazonal e W_t o ruído. No nosso caso, o tempo t está nos trimestres 1960.00, 1960.25, então uma unidade de tempo é um ano.

a) Ajuste o modelo de regressão

$$x_t = \beta t + \alpha_1 Q_1(t) + \alpha_2 Q_2(t) + \alpha_3 Q_3(t) + \alpha_4 Q_4(t) + W(t)$$

onde $Q_i(t) = 1$ se o tempo t corresponde ao trimestre $i = 1, 2, 3, 4$ e zero caso contrário. Os $Q_i(t)$ são chamados de variáveis indicadoras. Vamos assumir por enquanto que W_t é uma sequência de ruído branco gaussiano.

```
trimestre <- factor(cycle(jj))
summary(fit.jj <- lm(log(jj) ~ time(log(jj)) + trimestre + 0))
```

```
##
## Call:
## lm(formula = log(jj) ~ time(log(jj)) + trimestre + 0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## time(log(jj))  1.672e-01  2.259e-03   74.00  <2e-16 ***
## trimestre1    -3.283e+02  4.451e+00  -73.76  <2e-16 ***
## trimestre2    -3.282e+02  4.451e+00  -73.75  <2e-16 ***
```

```
## trimestre3    -3.282e+02  4.452e+00  -73.72   <2e-16 ***
## trimestre4    -3.284e+02  4.452e+00  -73.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic: 2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

b) Se o modelo estiver correto, qual é o aumento médio anual estimado do lucro por ação registrado?

O modelo ajustou um crescimento anual de 0.1672 para o retorno da ação. Voltando para a escala do lucro por ação, temos $e^{0.1672} = 1.182$, ou seja um crescimento de 18,2% anual.

c) Se o modelo estiver correto, a taxa de lucro média registrada aumenta ou diminui do terceiro para o quarto trimestre? E, por qual porcentagem aumenta ou diminui?

```
t_q3 <- exp(coefficients(fit.jj)[4])
t_q4 <- exp(coefficients(fit.jj)[5])
(t_q4-t_q3)/t_q3*100
```

```
## trimestre4
## -23.56716
```

A taxa de lucro média diminui em torno de 23,5% do terceiro para o quarto trimestre.

d) O que acontece se você incluir um termo de intercepto no modelo em (a)? Explique por que houve um problema.

```
summary(fit2.jj <- lm(log(jj) ~ time(log(jj)) + trimestre))
```

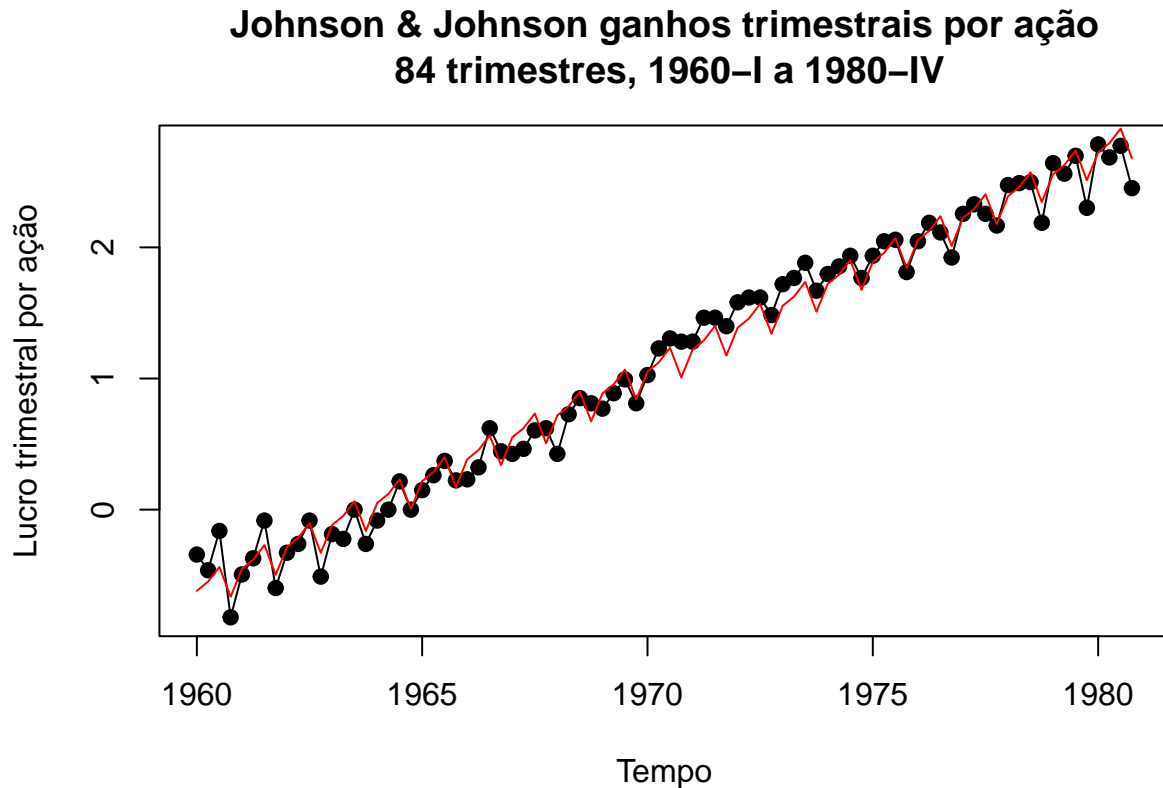
```
##
## Call:
## lm(formula = log(jj) ~ time(log(jj)) + trimestre)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.283e+02  4.451e+00 -73.761  < 2e-16 ***
## time(log(jj))  1.672e-01  2.259e-03  73.999  < 2e-16 ***
## trimestre2     2.812e-02  3.870e-02   0.727   0.4695
## trimestre3     9.823e-02  3.871e-02   2.538   0.0131 *
## trimestre4    -1.705e-01  3.873e-02  -4.403  3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic: 1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

Agora no modelo ajustado, o com intercepto representa o lucro médio do primeiro trimestre, enquanto os outros coeficientes indicam a diferença desse lucro médio do primeiro para o do primeiro trimestre. Portanto, o intercepto é equivalente ao valor estimado para trimestre1 no ajuste anterior.

e) Represente graficamente os dados X_t e sobreponha os valores ajustados, digamos $\hat{X}(t)$, no gráfico. Examine os resíduos $X_t - \hat{X}(t)$, e exponha suas conclusões. Parece que o modelo se ajusta bem aos

dados, ou seja, os resíduos parecem um ruído branco?

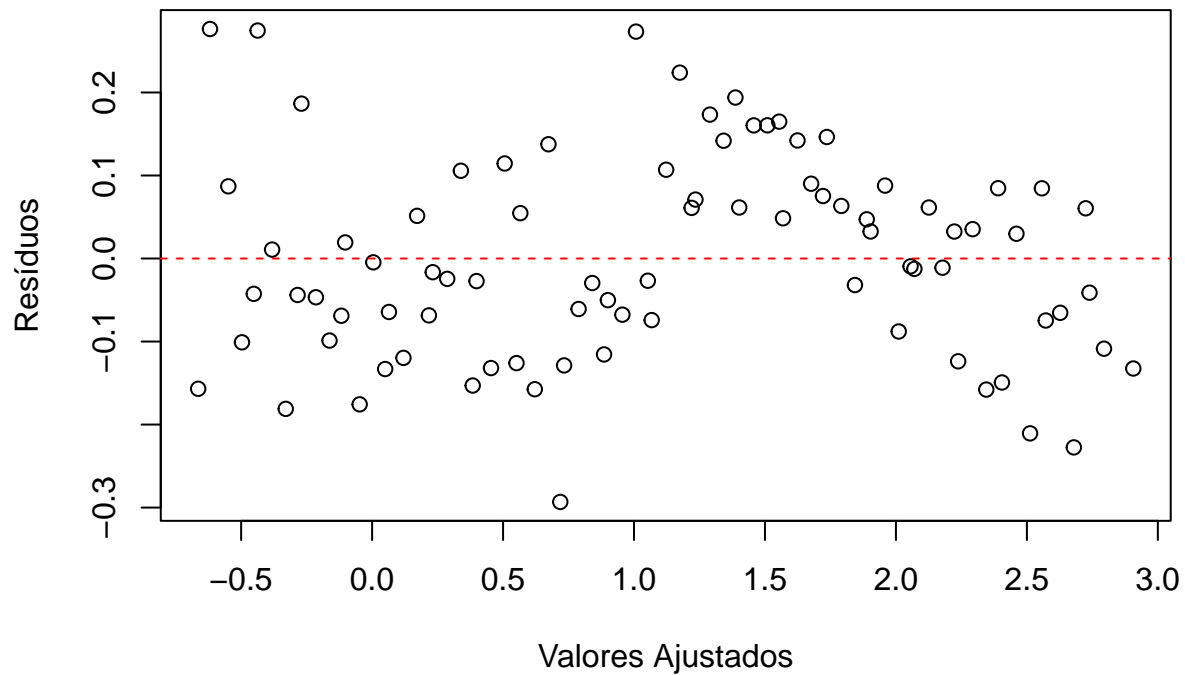
```
ajustados <- data.frame(x=time(jj), y =fitted(fit.jj))
residuos <- residuals(fit.jj)
plot(log(jj), type="o", xlab="Tempo", ylab="Lucro trimestral por ação", pch=19,
      main="Johnson & Johnson ganhos trimestrais por ação\n 84 trimestres, 1960-I a 1980-IV")
lines(ajustados, col="red")
```



O modelo aparentemente se ajusta bem aos dados tanto quanto a sua tendência quanto a seu comportamento sazonal dos resultados.

```
plot(fitted(fit.jj), residuos,
     xlab = "Valores Ajustados",
     ylab = "Resíduos",
     main = "Resíduos vs Valores Ajustados")
abline(h = 0, col = "red", lty = 2)
```

Resíduos vs Valores Ajustados



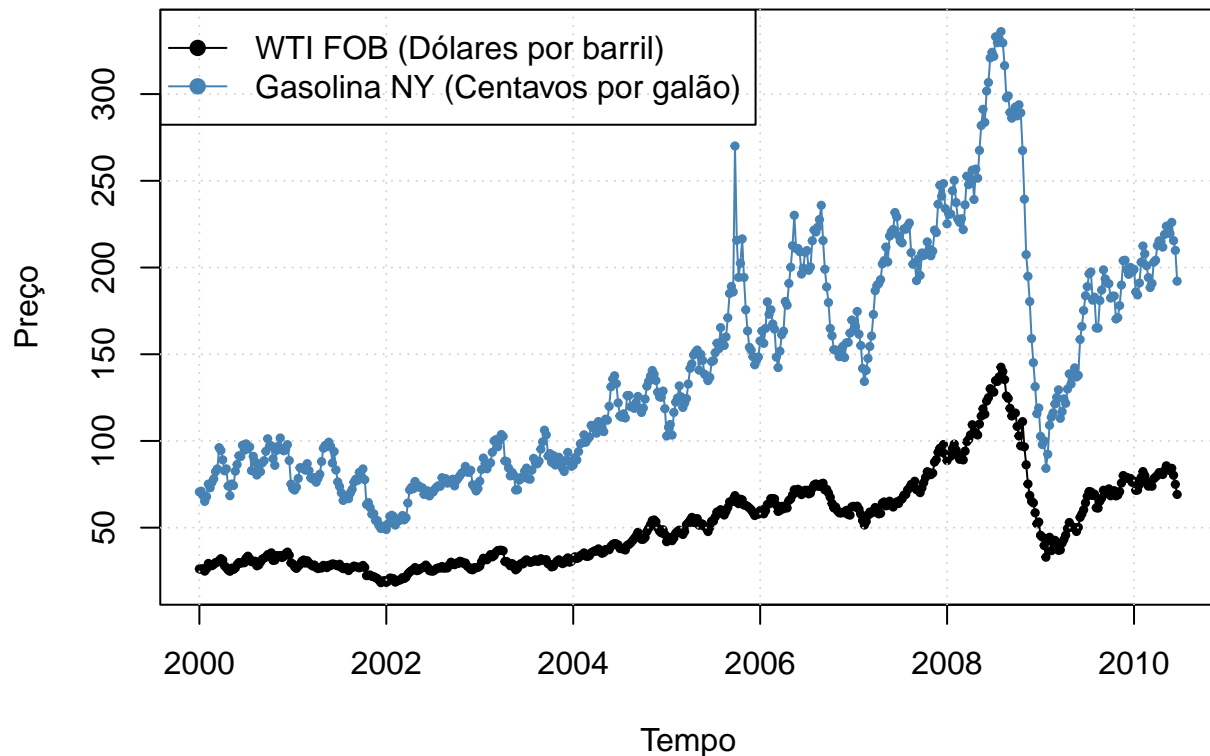
O resíduo aparente se comporta como um ruído branco e, portanto, o modelo se sustenta como bem ajustado para essa série temporal.

9- Considere as duas séries temporais de petróleo oil e gás gas, pacote astsa. A série de petróleo é em dólares por barril, enquanto a série de gás é em centavos por galão.

a) Mostre os dados no mesmo gráfico. Você acredita que as séries são estacionárias? Explique sua resposta.

```
par(mar=c(4,4,3,1))
plot(oil, type="o", xlab="Tempo", ylab="Preço", pch=19, cex=0.5, ylim=range(c(oil, gas)),
     main="Preço do WTI FOB e da gasolina (2000-1 até 2010-25)")
grid()
lines(gas, type="o", pch=19, cex=0.5, col="steelblue")
legend("topleft", legend=c("WTI FOB (Dólares por barril)", "Gasolina NY (Centavos por galão)"),
     col=c("black", "steelblue"), pch=19, lty=1)
```

Preço do WTI FOB e da gasolina (2000–1 até 2010–25)



Ambas as séries não parecem ser estacionárias, dado a sua tendência de aumento de preço, o que indicaria uma média não constante, e a grande queda abrupta observada em ambas, que pode ser indicativo de grande variância nesse intervalo. Esse comportamento é o contrário de uma série estacionária.

- b) Em economia, muitas vezes é a variação percentual no preço, denominada taxa de crescimento ou retorno, em vez da mudança absoluta de preço, que é importante. Argumente que uma transformação da forma $Y_t = \nabla \log(X_t)$ pode ser aplicada aos dados, onde X_t é a série de preços do petróleo ou gás.

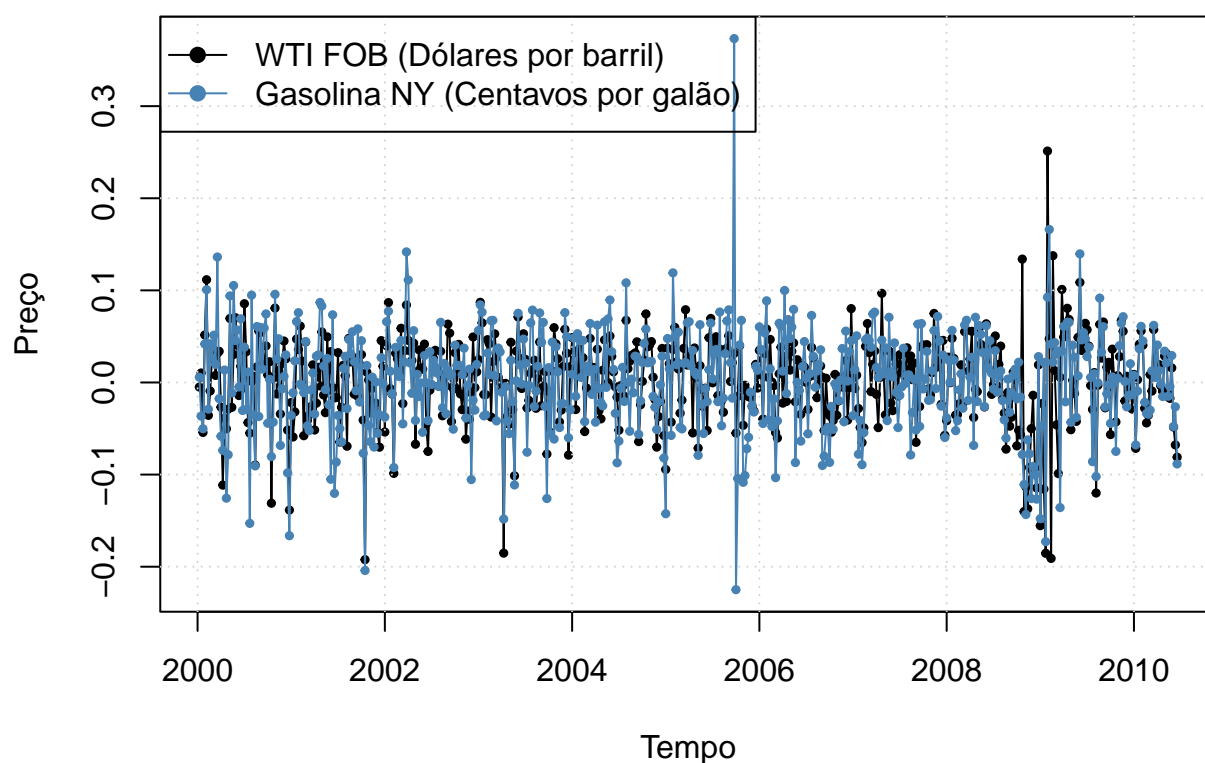
Essa transformação de retorno pode ser aplicada nos dados pois mostra a variação relativa nos preços ao longo do tempo, que é uma medida melhor do que a variância absoluta, ainda mais quando comparando séries de preços de produtos diferentes.

- c) Transforme os dados conforme descrito na parte (b), mostre os dados no mesmo gráfico, observe os exemplos de ACFs dos dados transformados e comente.

```
r_oil <- diff(log(oil))
r_gas <- diff(log(gas))

par(mar=c(4,4,3,1))
plot(r_oil, type="o", xlab="Tempo", ylab="Preço", pch=19, cex=0.5, ylim=range(c(r_oil, r_gas)),
     main="Preço do WTI FOB e da gasolina (2000-1 até 2010-25)")
grid()
lines(r_gas, type="o", pch=19, cex=0.5, col="steelblue")
legend("topleft", legend=c("WTI FOB (Dólares por barril)", "Gasolina NY (Centavos por galão)"),
      col=c("black", "steelblue"), pch=19, lty=1)
```

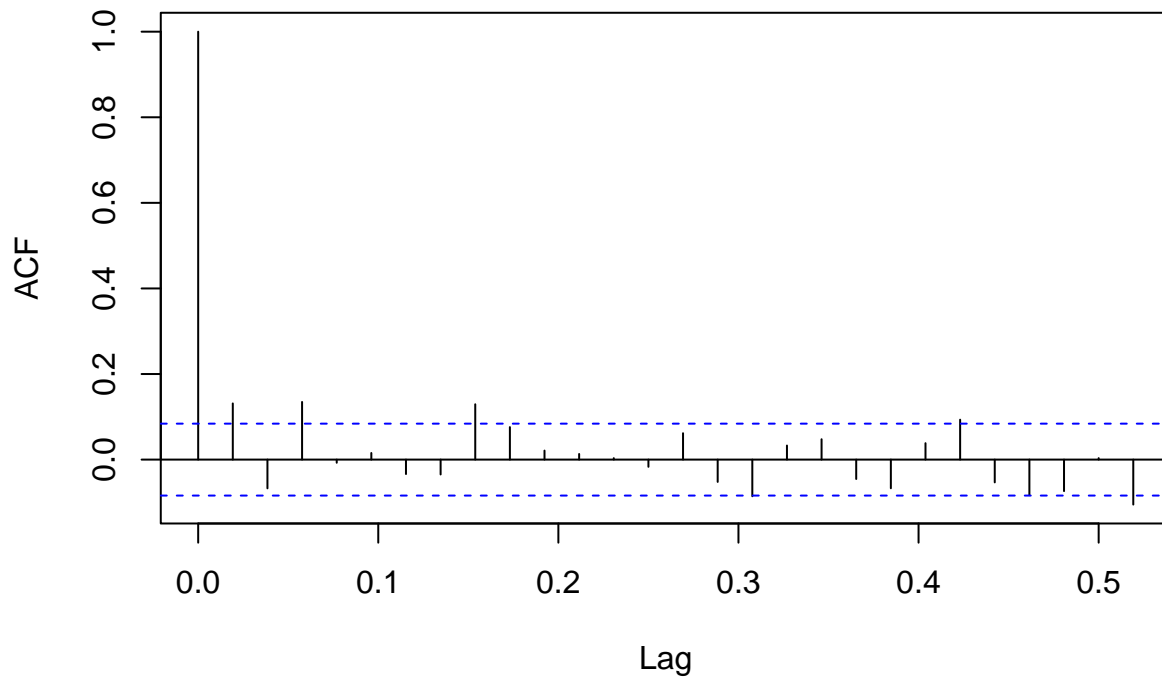
Preço do WTI FOB e da gasolina (2000–1 até 2010–25)



Agora sim as séries aparentar ser estacionárias, pois mantêm uma média constante ao redor de zero, além de uma variância também aparentemente constante. Vamos continuar a análise olhando a ACF amostral.

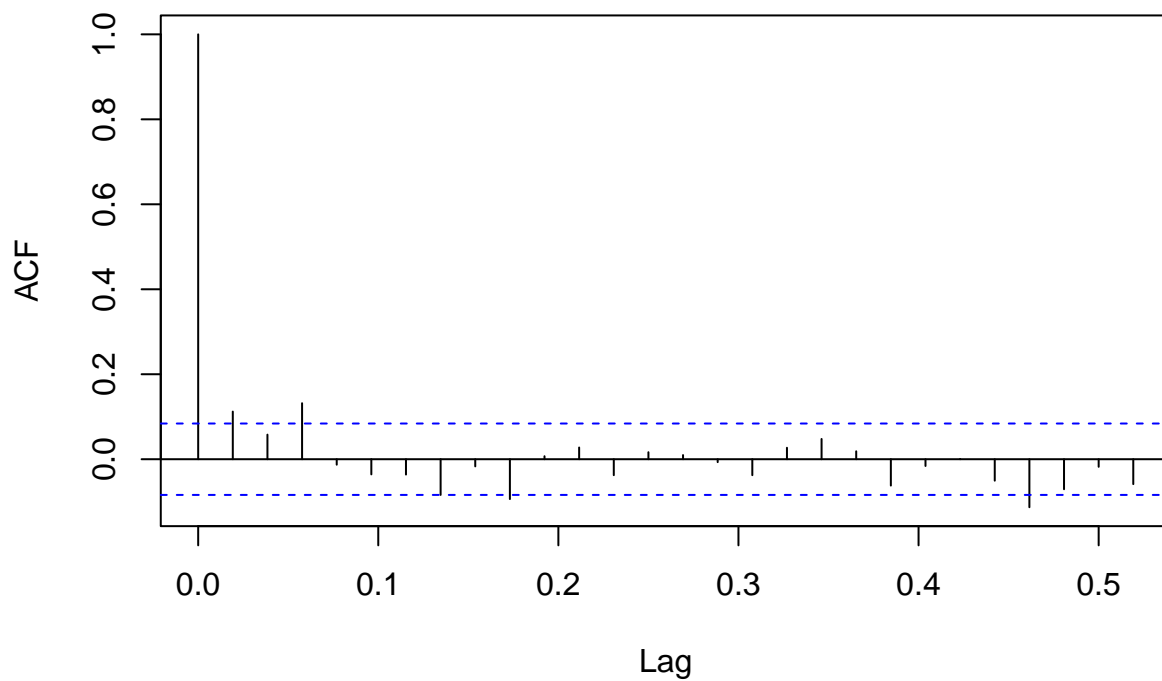
```
acf(r_oil)
```

Series r_oil



```
acf(r_gas)
```

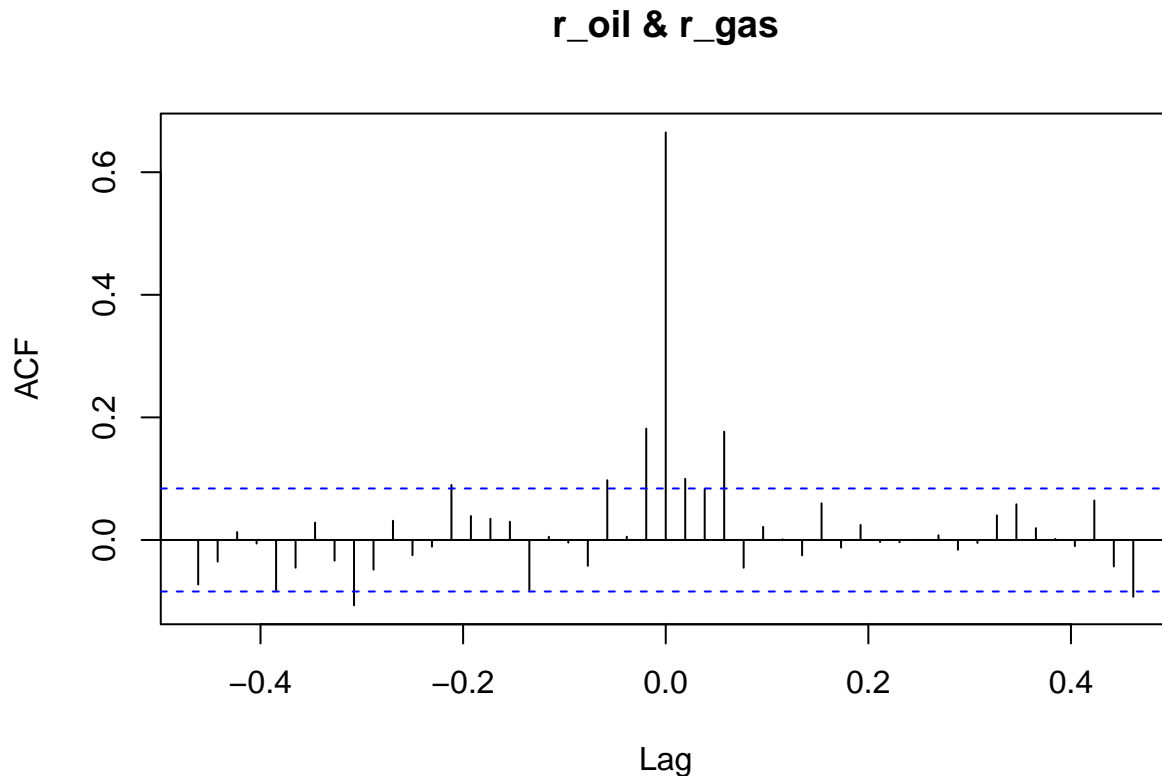
Series r_gas



A ACF amostral $\hat{\rho}(h)$ indica que a maior parte das autocorrelações está dentro do intervalo de confiança, exceto quando $h = 0$. Isso sugere que ambas as séries de retorno parecem ser estacionárias, sem padrões de autocorrelação significativa para $h \geq 1$.

- d) Grafique o CCF dos dados transformados e comente. Os valores pequenos, mas significativos, quando o gás conduz o petróleo podem ser considerados como explicativos.

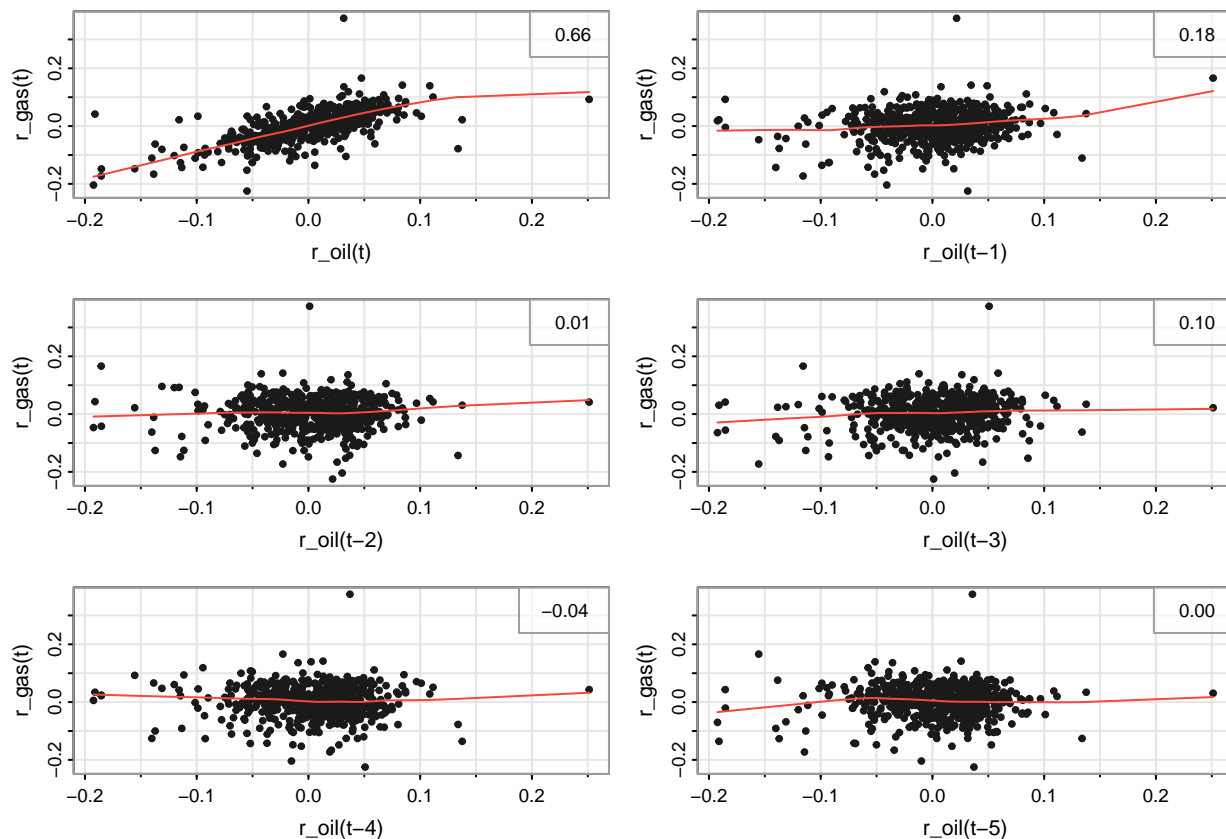
```
ccf(r_oil, r_gas)
```



A análise de correlação cruzada entre as séries temporais tem valores significativos para valores próximos de zero, o que nos leva a conclusão de que os preços semanais de uma série influencia a outra, principalmente na mesma semana. Como os valores significativos se estendem mais à direita, essa é uma evidência de que a série de preços no petróleo acaba por conduzir o preço do combustível até três semanas à frente.

- e) Exiba gráficos de dispersão da série de taxas de crescimento de petróleo e gás por até três semanas do prazo de entrega dos preços do petróleo; inclua um suavizador não paramétrico em cada gráfico e comente os resultados, por exemplo, existem outliers? As relações são lineares?

```
lag2.plot(r_oil, r_gas, 5, cex=0.6, pch=19)
```

Os gráficos acima mostram uma relação não linear para os instantes $t, t-1, t-3$. A correlação diminui à medida que o lag aumenta, sugerindo que a dependência entre as séries é mais forte no curto prazo e se dissipa rapidamente.

```
dates <- time(r_oil)
outliers_oil <- which(abs(r_oil - mean(r_oil)) > 3 * sd(r_oil))
outliers_gas <- which(abs(r_gas - mean(r_gas)) > 3 * sd(r_gas))
outliers <- unique(c(outliers_oil, outliers_gas))

outliers_df <- data.frame(
  Date = dates[outliers],
  r_oil = r_oil[outliers],
  r_gas = r_gas[outliers])
outliers_df[order(outliers_df$Date), ]
```

```
##      Date      r_oil      r_gas
## 1 2001.788 -0.19233834 -0.20407835
## 2 2003.269 -0.18527189 -0.14825319
## 8 2005.731  0.03160266  0.37335871
## 9 2005.750 -0.05493578 -0.22494557
## 3 2008.827 -0.14013548 -0.11074274
## 4 2009.000 -0.15546560 -0.14782088
## 5 2009.058 -0.18544987 -0.17281434
## 6 2009.077  0.25124705  0.09252597
## 7 2009.115 -0.19099605  0.04156414
```

Os outliers foram definidos arbitrariamente como valores que estão a mais de 3 desvios padrão da média. Ao analisar as datas dos outliers identificados, observa-se que elas coincidem com períodos de instabilidade econômica. Por exemplo, a crise financeira de 2008, e suas consequências no ano de 2009, é coincidente com

cinco dos outliers. Outras datas, como 2001 e 2003, também correspondem a momentos críticos na economia global, como a bolha da internet e a invasão do Iraque, que impactaram os mercados de petróleo e gás.

f) Tem havido uma série de estudos questionando se os preços da gasolina respondem mais rapidamente quando os preços do petróleo estão subindo do que quando os preços do petróleo estão caindo, chamamos esse efeito de assimetria dos preços. Vamos tentar explorar essa questão aqui com regressão retardada simples. Ignorando alguns problemas óbvios, como outliers e erros autocorrelacionados, por isso não será uma análise definitiva. Considere G_t e O_t denotar as taxas de crescimento dos preços do gás e de petróleo.

g) Ajuste o seguinte modelo de regressão e comente os resultados: $G_t = \alpha_1 + \alpha_2 I_t + \beta_1 O_t + \beta_2 O_{t-1} + W_t$, onde $I_t = 1$ se $O_t \geq 0$ e 0 (zero) caso contrário, ou seja, I_t é o indicador de ausência de crescimento ou crescimento positivo no preço do petróleo.

```
poil = diff(log(oil))
pgas = diff(log(gas))
indi = ifelse(poil < 0, 0, 1)
mess = ts.intersect(pgas, poil, poilL = lag(poil,-1), indi)
summary(fit <- lm(pgas ~ poil + poilL + indi, data=mess))

##
## Call:
## lm(formula = pgas ~ poil + poilL + indi, data = mess)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18451 -0.02161 -0.00038  0.02176  0.34342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006445   0.003464  -1.860  0.06338 .
## poil        0.683127   0.058369  11.704 < 2e-16 ***
## poilL       0.111927   0.038554   2.903  0.00385 **
## indi        0.012368   0.005516   2.242  0.02534 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04169 on 539 degrees of freedom
## Multiple R-squared:  0.4563, Adjusted R-squared:  0.4532
## F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

ii. Qual é o modelo ajustado quando há um crescimento negativo no preço do petróleo no tempo t ? Qual é o modelo ajustado quando não há crescimento positivo do preço do petróleo? Esses resultados suportam a hipótese de assimetria?

O modelo é dado por $G_t = \alpha_1 + \alpha_2 + \beta_1 O_t + \beta_2 O_{t-1} + W_t$. Quando há crescimento positivo, e substituindo os valores encontrados, temos $G_t = -0.0064 + 0.6831 O_t + 0.1119 O_{t-1} + W_t$.

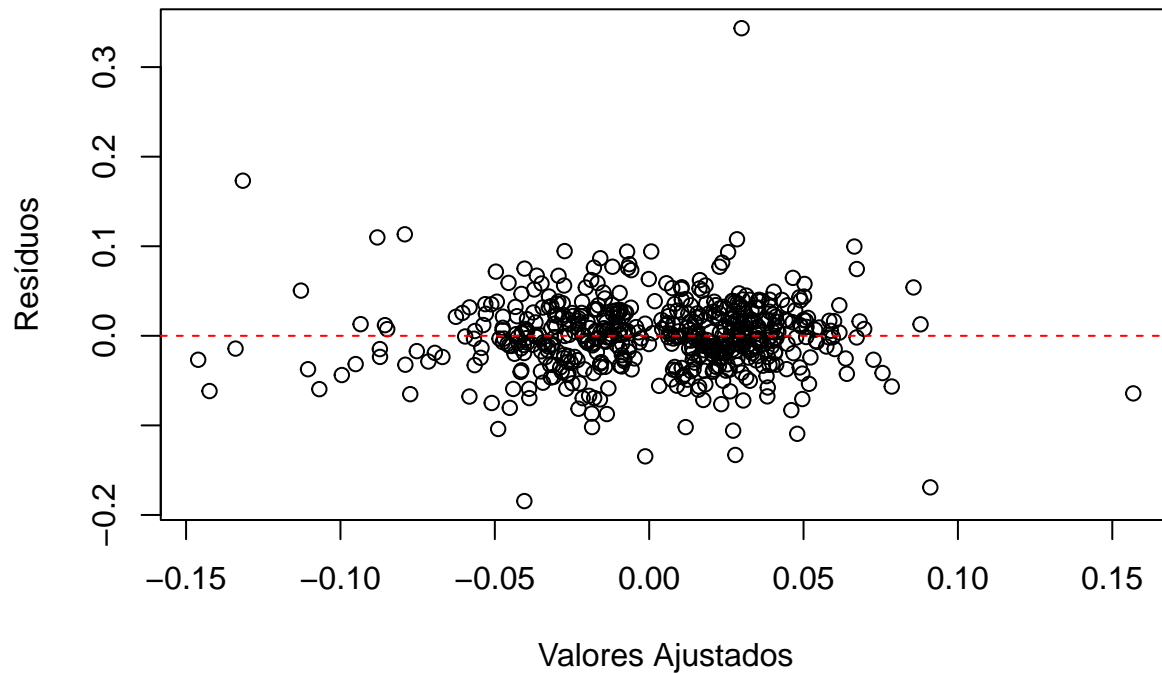
Já quando há crescimento negativo, o modelo fica $G_t = -0.0064 + 0.0124 + 0.6831 O_t + 0.1119 O_{t-1} + W_t = 0.006 + 0.6831 O_t + 0.1119 O_{t-1} + W_t$. Considerando essa diferença de sinal dos β 's estimados (-0.0064 vs 0.006), esses resultados indicam que o impacto dos preços do petróleo sobre os preços da gasolina é assimétrico. Em situações de crescimento positivo do petróleo, a resposta da gasolina é expressivamente maior comparado ao que ocorre quando há crescimento negativo.

iii. Analise os resíduos do ajuste e comente.

```
ajustados <- data.frame(x = time(r_oil)[-1], y = fitted(fit))
residuos <- residuals(fit)
```

```
plot(fitted(fit), residuos,
     xlab = "Valores Ajustados",
     ylab = "Resíduos",
     main = "Resíduos vs Valores Ajustados")
abline(h = 0, col = "red", lty = 2)
```

Resíduos vs Valores Ajustados



Os resíduos do ajuste se comportam de maneira adequada em torno de zero, sem padrões constantes de fuga da média constatados. Portanto, o modelo se sustenta como explicativo às séries temporais e também ao seu comportamento da série de preços do petróleo conduzir a série de preços da gasolina, inclusive de forma assimétrica, sendo mais alongada temporalmente quando o crescimento de preços é positivo.