

Mais sobre modelagem

Prof. Caio Azevedo

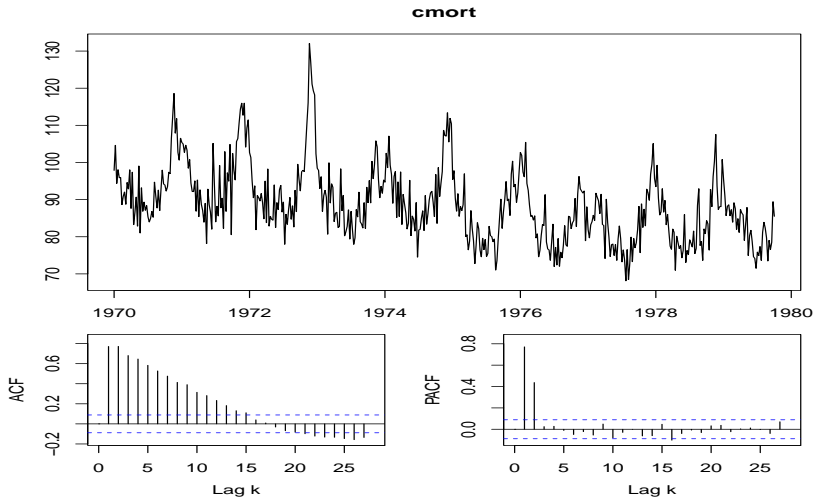
Introdução

- Discutiremos um pouco mais sobre modelagem de ST, limitando-nos a classe de modelos **ARIMA**. Ou seja, mesmo o aqui escolhido pode ou não ser apropriado, ou mesmo pode não ser o mais apropriado.
- Essencialmente usaremos as metodologias vistas (gráficos de ST, FAC, FACP, eventualmente suavização, estimação por MV, análise de resíduos padronizados e previsão) usando a função `sarima` do pacote `astsa`.
- Revisitaremos algumas séries vistas ao longo do curso, até o momento.

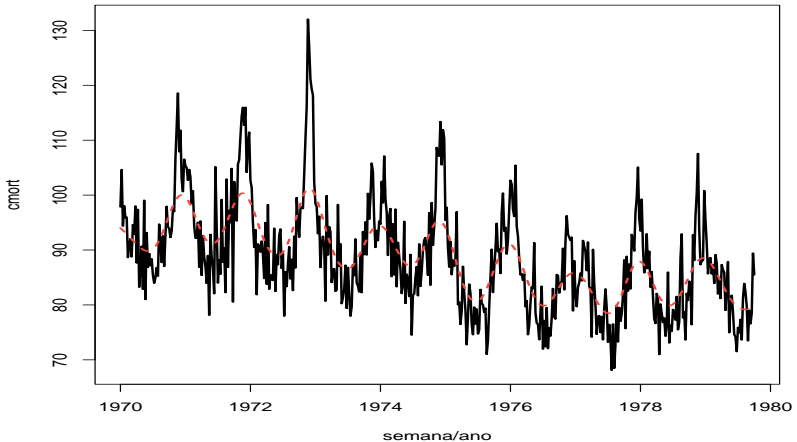
Mortalidade por questões cardiovasculares

- Mortalidade média semanal por questões cardiovasculares, no condado de Los Angeles, no período de 1970-1979.
- Disponível no pacote `astsa` sob o nome “`cmort`”.
- A ST parece ser estacionária (com sazonalidade).
- Um modelo $AR(p)$ parece ser apropriado.
- Exercício: escrever os modelos ajustados.

Gráficos de ST, FAC e FACP da ST

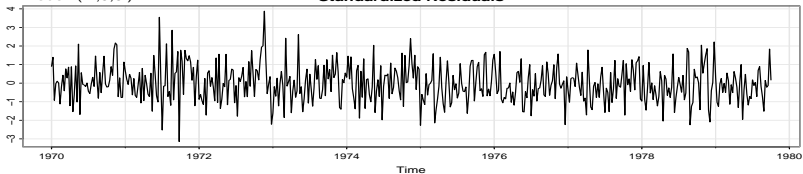


Suavização por kernel ($b=0,5$)

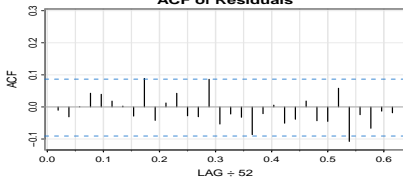


Model: (2,0,0)

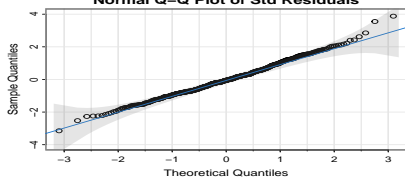
Standardized Residuals



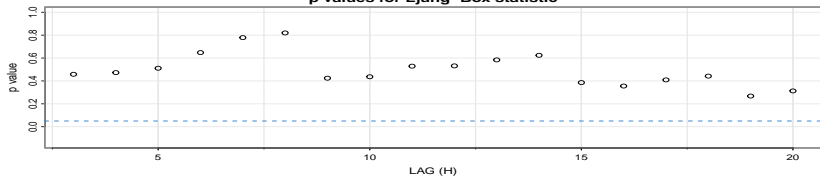
ACF of Residuals

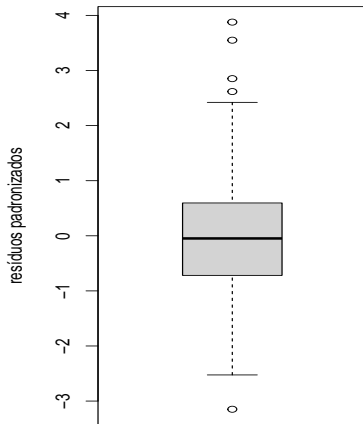
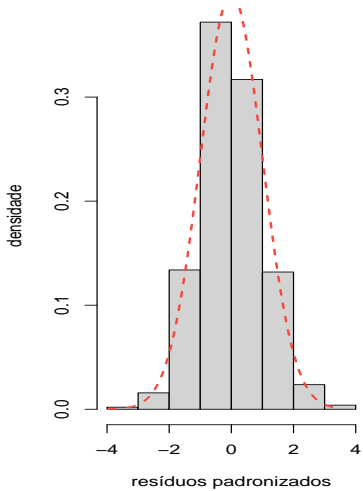


Normal Q-Q Plot of Std Residuals

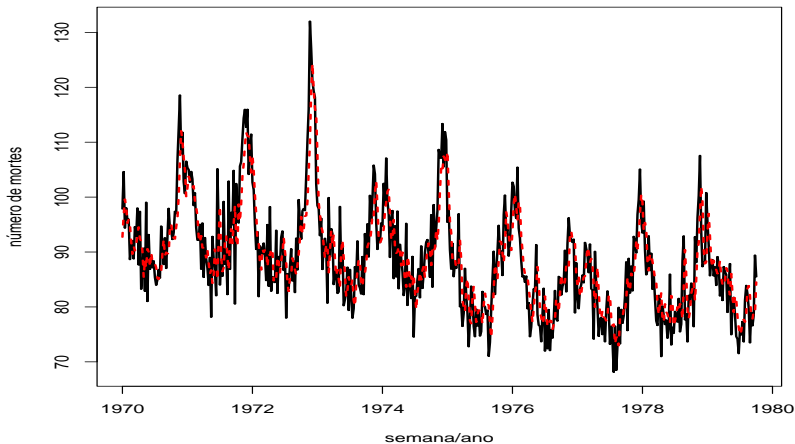


p values for Ljung-Box statistic

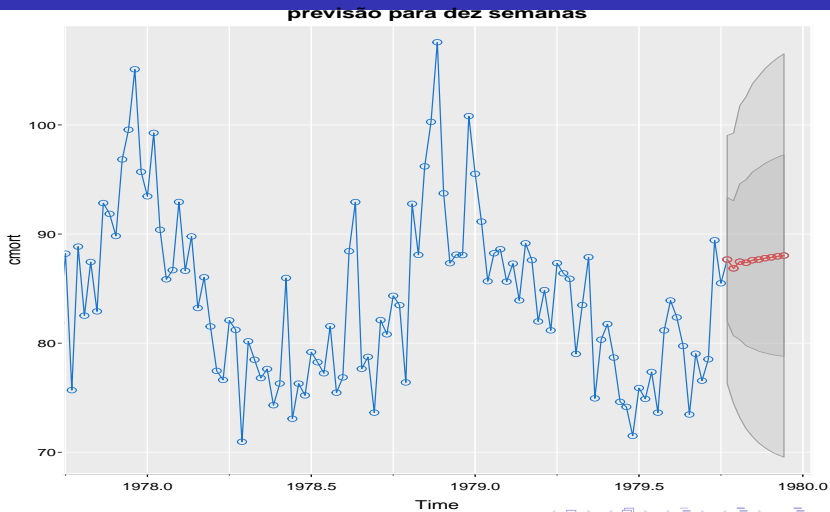




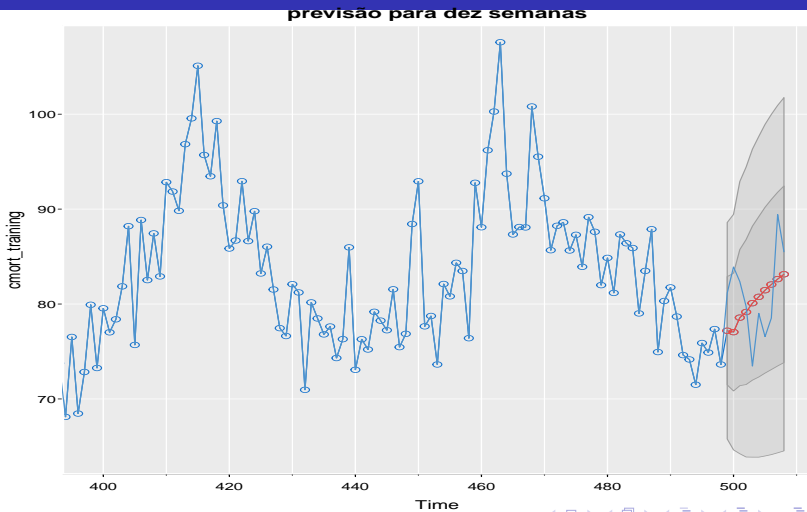
Previsão para valores observado vermelho (P), preto (P)



Previsão para valores futuros

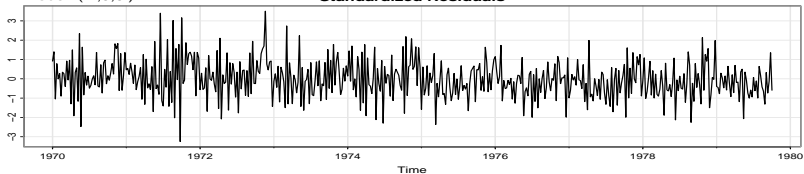


Previsão para valores futuros observados

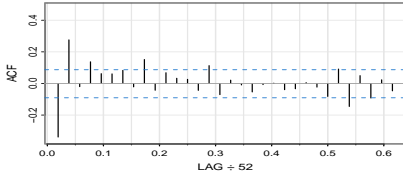


Model: (1,0,0)

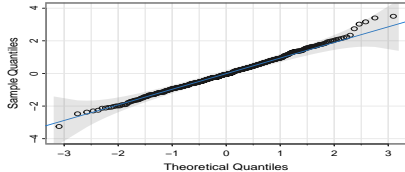
Standardized Residuals



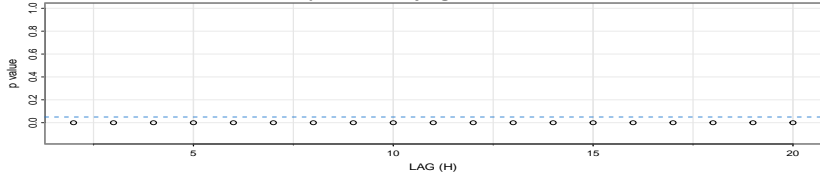
ACF of Residuals

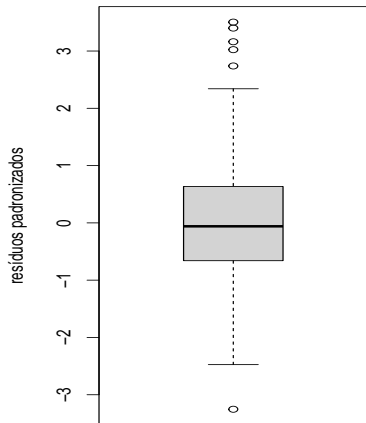
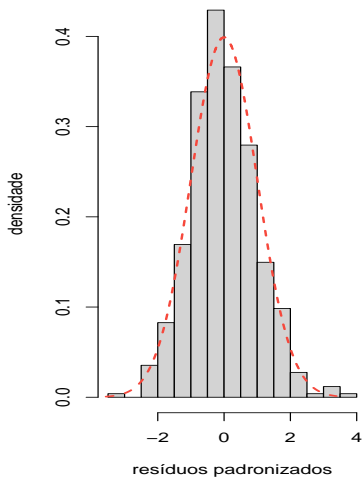


Normal Q-Q Plot of Std Residuals



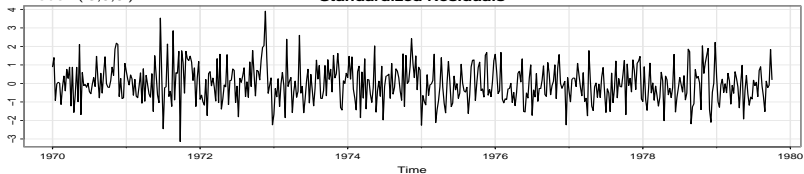
p values for Ljung-Box statistic



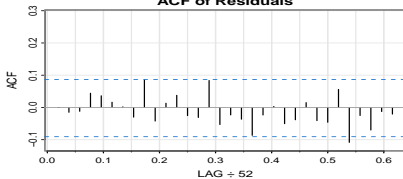


Model: (3,0,0)

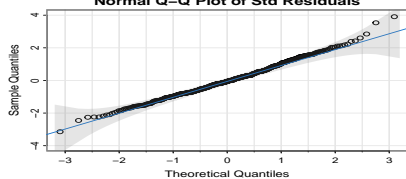
Standardized Residuals



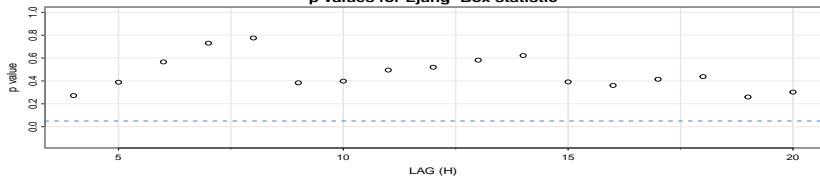
ACF of Residuals

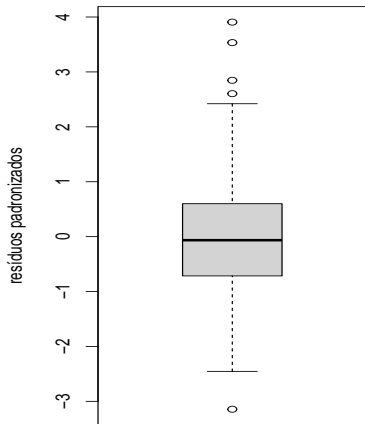
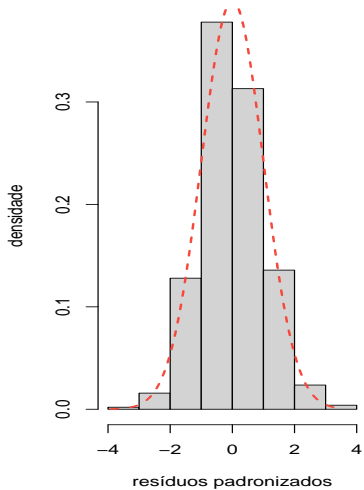


Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic





Ajuste dos modelos

| Modelo | AIC | AICc | BIC |
|--------|-------|-------|-------|
| AR(1) | 6,547 | 6,547 | 6,572 |
| AR(2) | 6,334 | 6,334 | 6,367 |
| AR(3) | 6,337 | 6,337 | 6,378 |

Previsão para valores observados

| Modelo | EQM | EAM | EPM | EPAM |
|--------|--------|-------|--------|-------|
| AR(1) | 40,268 | 5,001 | -0,520 | 5,643 |
| AR(2) | 32,371 | 4,494 | -0,445 | 5,061 |
| AR(3) | 32,350 | 4,493 | -0,448 | 5,059 |

Previsão para valores futuros

| Modelo | EQM | EAM | EPM | EPAM |
|--------|--------|-------|--------|-------|
| AR(1) | 51,079 | 5,851 | -7,639 | 7,639 |
| AR(2) | 18,393 | 3,547 | -2,737 | 4,615 |
| AR(3) | 17,463 | 3,478 | -2,504 | 4,520 |

Estimativas

| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|----------|-------|------|---------------|----------|---------|
| ϕ_1 | 0,43 | 0,04 | 0,35 ; 0,51 | 10,84 | <0,0001 |
| ϕ_2 | 0,44 | 0,04 | 0,36 ; 0,52 | 11,12 | <0,0001 |
| μ | 88,85 | 1,94 | 85,05 ; 92,66 | 45,78 | <0,0001 |

raízes de $\phi(z) \approx (1,094; 2,066)$ (modelo estacionário e causal)

| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|----------|-------|------|-----------------|----------|---------|
| ϕ_1 | 0,42 | 0,04 | [0,33 ; 0,51] | 9,47 | <0,0001 |
| ϕ_2 | 0,43 | 0,04 | [0,34 ; 0,52] | 9,76 | <0,0001 |
| ϕ_3 | 0,03 | 0,04 | [-0,06 ; 0,11] | 0,57 | 0,5673 |
| μ | 88,87 | 1,99 | [84,98 ; 92,77] | 44,70 | <0,0001 |

Comentários

- Os modelos se ajustou AR(2) e AR(3) se ajustaram bem aos dados (análise residual + análise preditiva para os valores observados e futuros), em que pese uma possível assimetria e/ou caudas pesadas, dos resíduos padronizados.
- O modelo AR(2) se ajustou melhor do que os demais, segundo os critérios de informação, enquanto que o AR(3) foi superior em relação aos critérios de previsão.

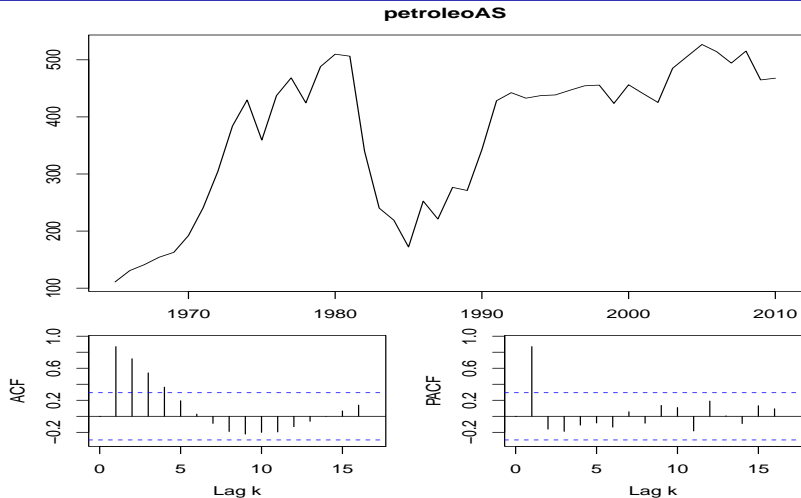
Comentários

- Mesmo com a não significância do ϕ_3 (modelo AR(3)), ficaríamos com esse modelo, sob ressalvas.
- Possibilidades de melhora (não exaustivas):
 - Considerar outra distribuição para os erros.
 - Ajustar um modelo SARIMA (com sazonalidade + AR, sem a componente MA) com a mesma ou outra distribuição para os erros.

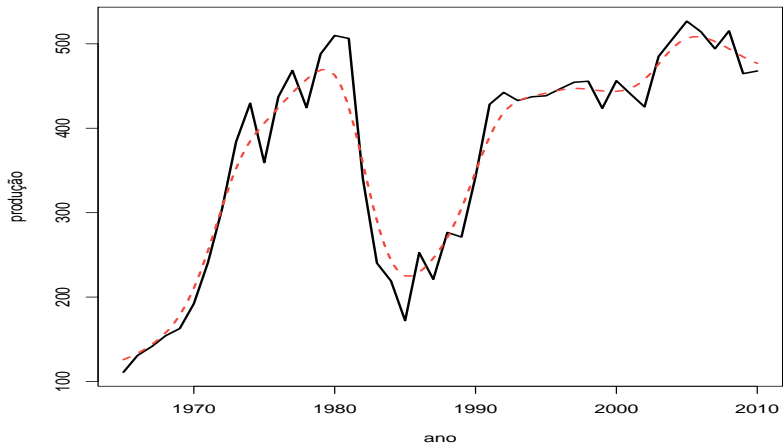
Produção de petróleo da Arábia Saudita

- Já mencionada anteriormente e inicialmente analisada ([aqui](#)).
- Produção anual de petróleo (milhões de toneladas), Arábia Saudita, 1965-2010. Um total de $n = 46$ observações.
- ST aparentemente estacionária modelável por um processo AR(1).
- Exercício: escrever os modelos ajustados.

Gráficos de ST, FAC e FACP da ST

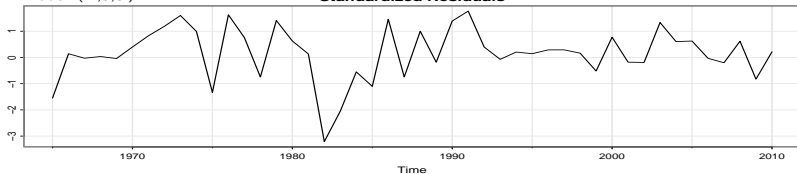


Suavização por kernel ($b=4$)

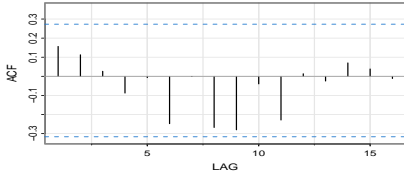


Model: (1,0,0)

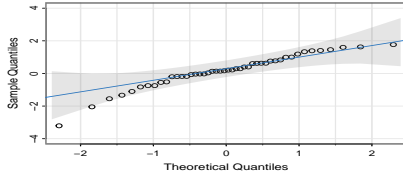
Standardized Residuals



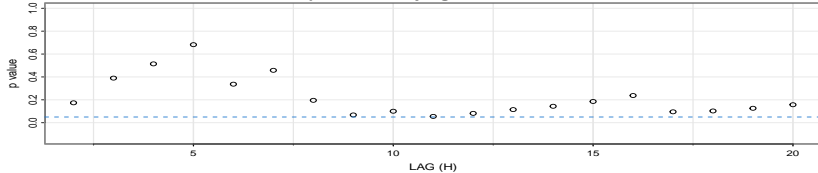
ACF of Residuals

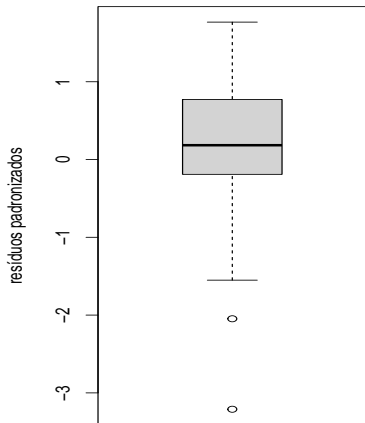
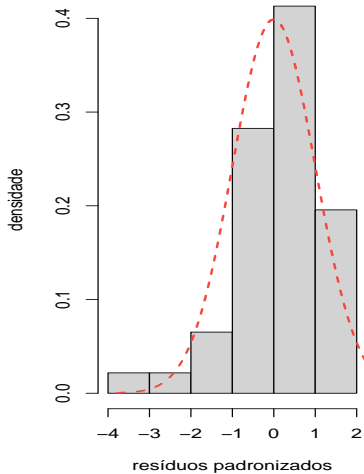


Normal Q-Q Plot of Std Residuals



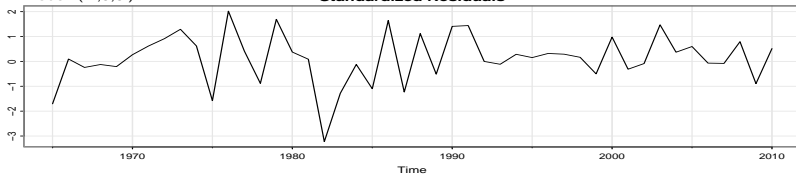
p values for Ljung-Box statistic



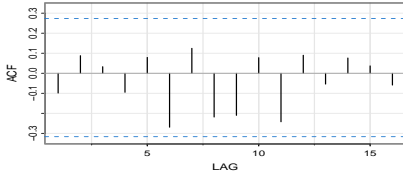


Model: (2,0,0)

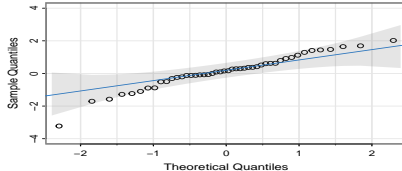
Standardized Residuals



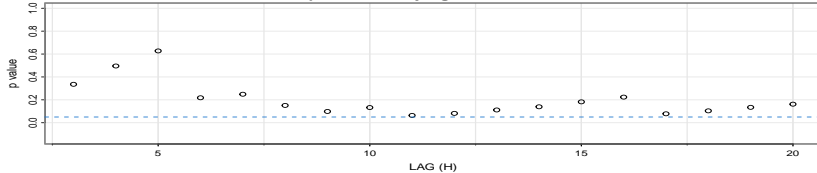
ACF of Residuals

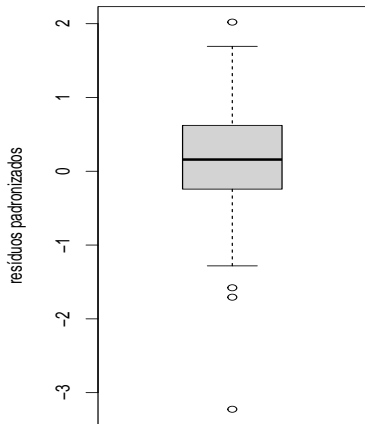
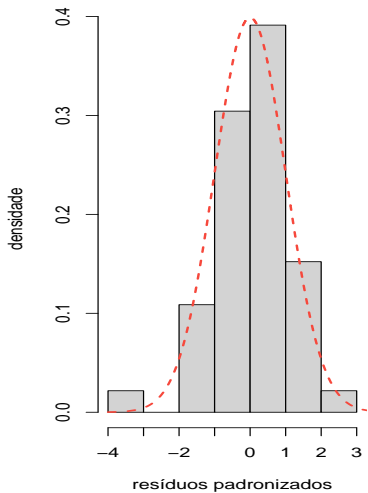


Normal Q-Q Plot of Std Residuals



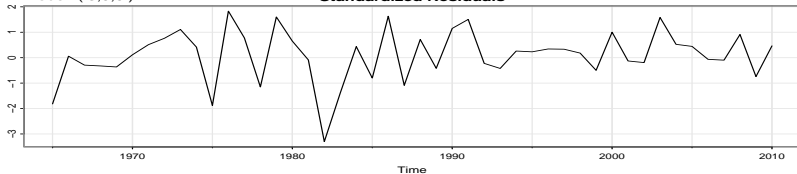
p values for Ljung-Box statistic



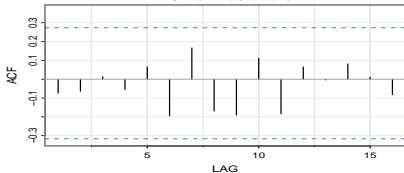


Model: (3,0,0)

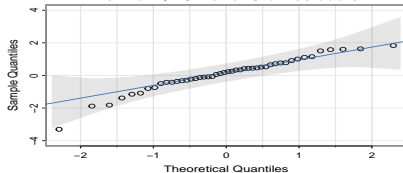
Standardized Residuals



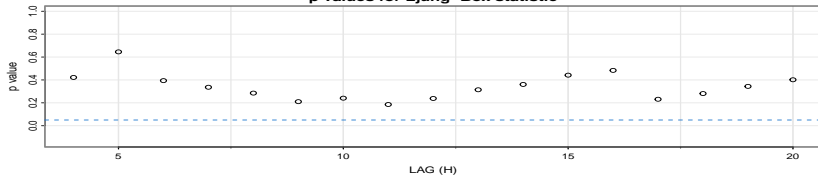
ACF of Residuals

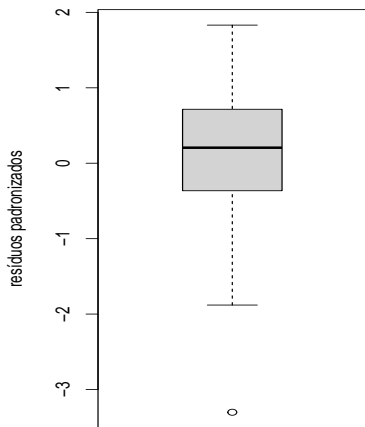
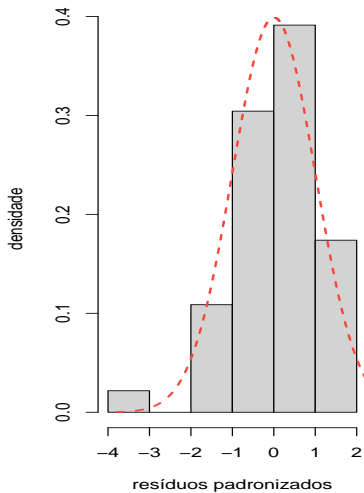


Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic





Estatísticas de Comparação de modelos

| Modelo | AIC | AICc | BIC |
|--------|--------|--------|--------|
| AR1 | 10,788 | 10,794 | 10,907 |
| AR2 | 10,775 | 10,788 | 10,934 |
| AR3 | 10,785 | 10,806 | 10,984 |

Ajustes dos modelos

| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|--------|--------|-------|-------------------|----------|---------|
| ϕ | 0,94 | 0,05 | [0,85 ; 1,04] | 19,55 | <0,0001 |
| μ | 336,97 | 96,91 | [147,04 ; 526,90] | 3,48 | 0,0012 |

raiz de $\phi(z)$: $|z| \approx 1,061$ (modelo estacionário e causal)

| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|----------|--------|-------|-------------------|----------|---------|
| ϕ_1 | 1,16 | 0,14 | [0,88 ; 1,44] | 8,21 | <0,0001 |
| ϕ_2 | -0,24 | 0,15 | [-0,53 ; 0,05] | -1,63 | 0,1100 |
| μ | 347,52 | 76,38 | [197,82 ; 497,22] | 4,55 | <0,0001 |

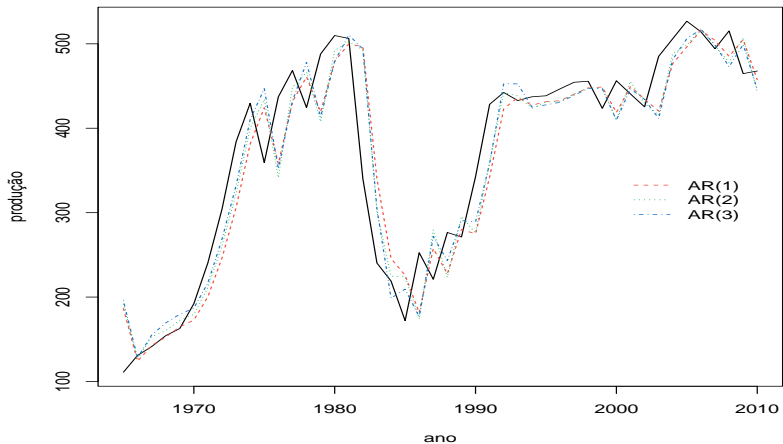
raízes de $\phi(z) \approx (1,122; 3,723)$ (modelo estacionário e causal)

Ajustes dos modelos

| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|----------|--------|-------|-------------------|----------|---------|
| ϕ_1 | 1,11 | 0,14 | [0,83 ; 1,39] | 7,80 | <0,0001 |
| ϕ_2 | -0,03 | 0,22 | [-0,46 ; 0,40] | -0,13 | 0,8995 |
| ϕ_3 | -0,19 | 0,15 | [-0,48 ; 0,11] | -1,26 | 0,2147 |
| μ | 353,41 | 61,21 | [233,45 ; 473,38] | 5,77 | <0,0001 |

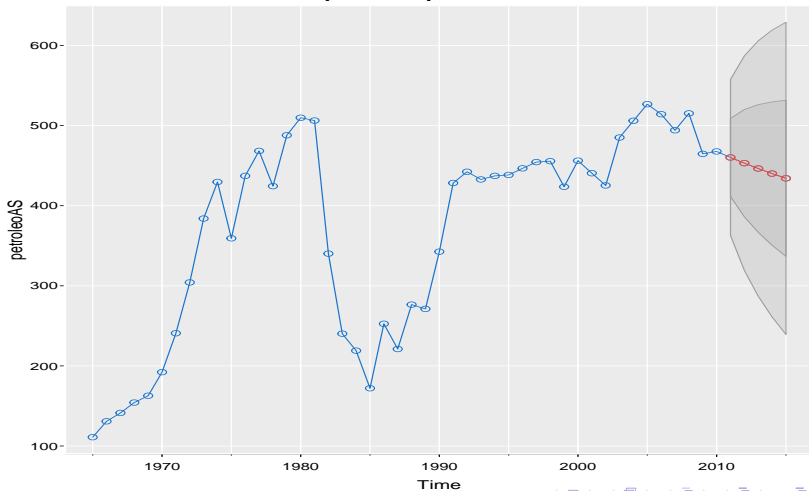
raízes de $\phi(z) \approx (1,359; 1,359; 2,854)$ (modelo estacionário e causal)

Previsão para valores observados



Previsão para valores futuros AR(1)

previsão para 5 anos



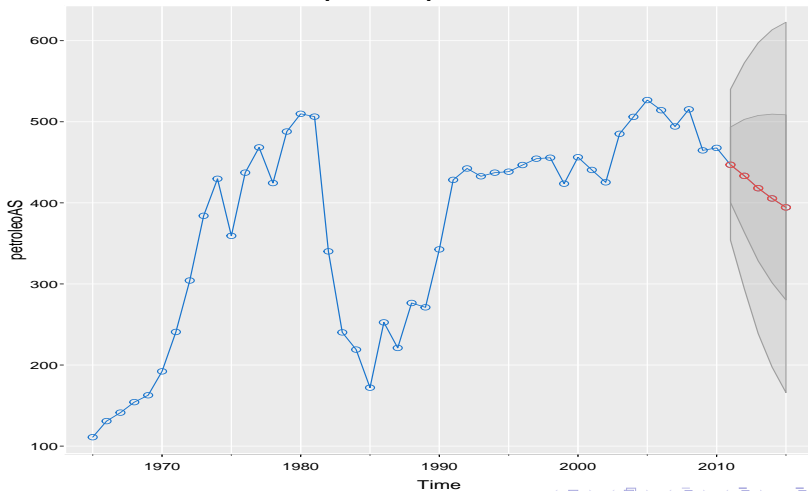
Previsão para valores futuros AR(2)

previsão para 5 anos

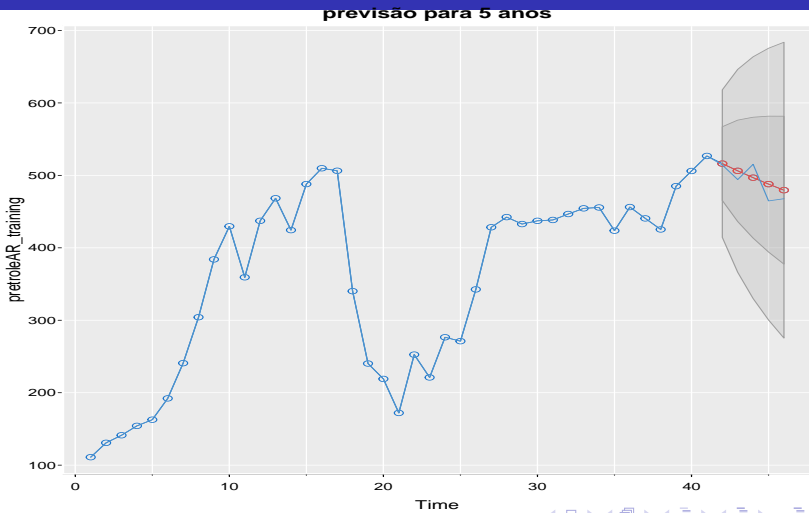


Previsão para valores futuros AR(3)

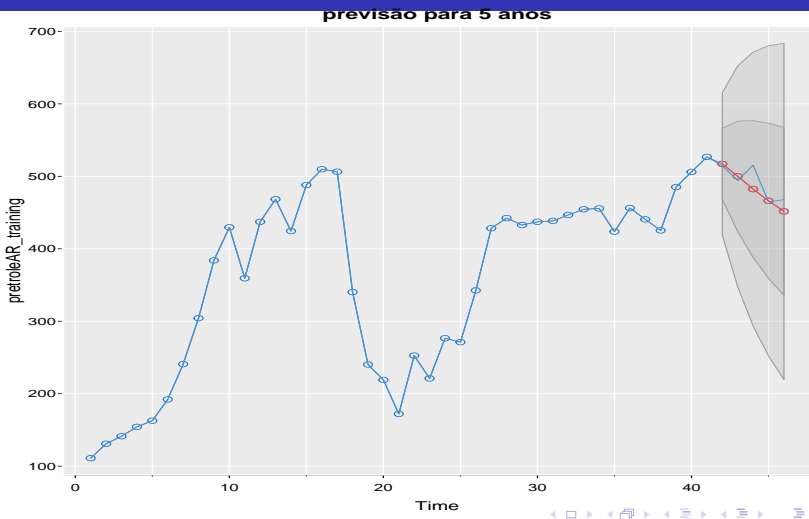
previsão para 5 anos



Previsão para valores futuros observados AR(1)

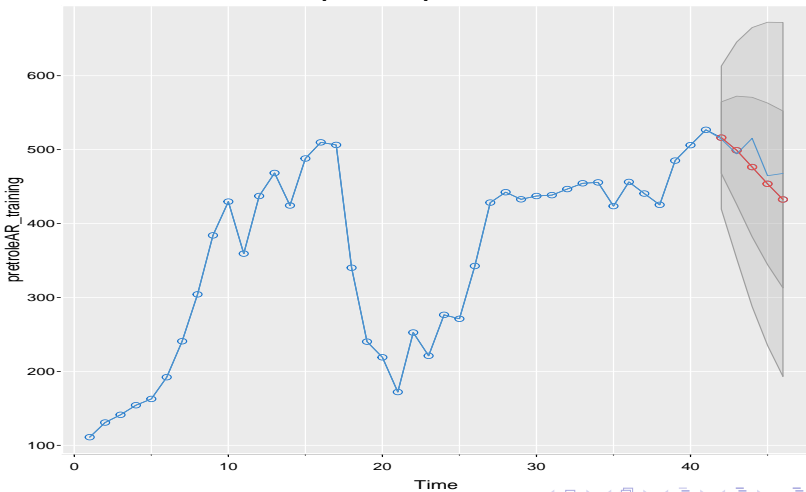


Previsão para valores futuros observados AR(2)



Previsão para valores futuros observados AR(3)

previsão para 5 anos



Previsão para valores observados

| Modelo | EQM | EAM | EPM | EPAM |
|--------|----------|--------|--------|--------|
| AR(1) | 2373,497 | 36,606 | 0,100 | 11,820 |
| AR(2) | 2242,634 | 35,243 | -0,685 | 11,562 |
| AR(3) | 2166,302 | 35,257 | -1,200 | 11,518 |

Previsão para valores futuros

| Modelo | EQM | EAM | EPM | EPAM |
|--------|----------|--------|--------|--------|
| AR(1) | 3404,386 | 56,214 | 11,065 | 11,065 |
| AR(2) | 4391,975 | 64,807 | 12,811 | 12,811 |
| AR(3) | 7148,534 | 83,427 | 16,552 | 16,552 |

Comentários

- Nenhum dos três modelos se ajustou bem aos dados: 1) Assimetria negativa dos resíduos e comportamento não compatível com um ruído branco, 2) Previsões ruins para valores observados, embora razoáveis para valores futuros. O modelo AR(3) apresentou o “melhor” ajuste.
- Provavelmente por isso, as estatísticas de comparação de modelo apresentaram resultado contraditório com as estimativas. Enquanto que a primeira indica o modelo AR(2), a segunda indica o modelo AR(1).

Comentários

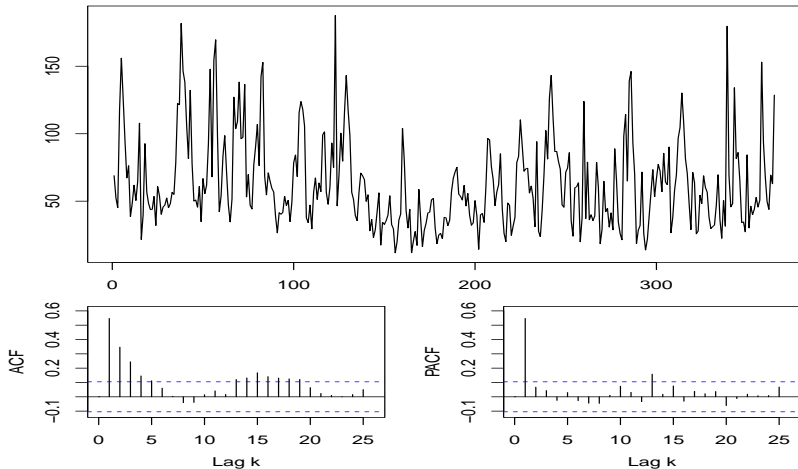
- Além disso, somente para previsão de valores futuros, as medidas de comparação concordam e, ainda assim, os erros podem ser não razoáveis.
- Sugestão: ajustar um modelo skew-t ou skew-normal com estrutura (AR(1), AR(2), AR(3)).

Concentração de Ozônio na cidade de São Paulo

- Trata-se da mensuração diária da concentração de Ozônio na atmosfera da cidade de São Paulo, durante ($n =$) 365 dias do ano de 1997.
- Unidade de medida (g/m^3)(?).
- Os efeitos para a saúde do ozônio (concentração acima do desejado) são parecidos com os de outros poluentes: tosse, irritação na garganta, nos olhos e no nariz. Em casos extremos, segundo os médicos, a poluição pode agravar problemas cardiovasculares e levar até ao infarto.
- A série aparente ser estacionária, com a eventual presença de sazonalidade e/ou ciclos. Um modelo AR(p) pode ser um opção, ao menos inicial, apropriada.

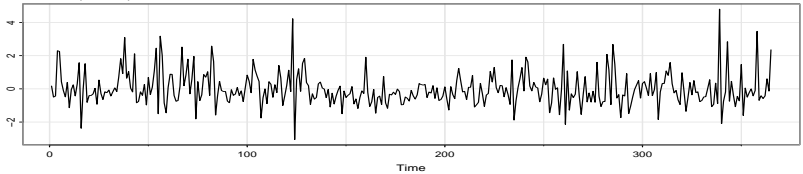
Gráficos de ST, FAC e FACP da ST

o3

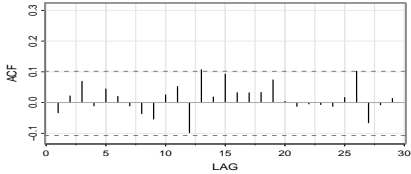


Model: (1,0,0)

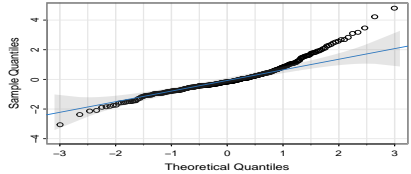
Standardized Residuals



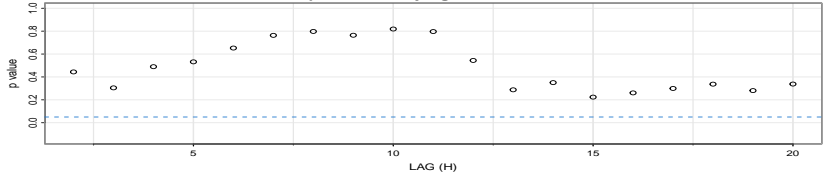
ACF of Residuals

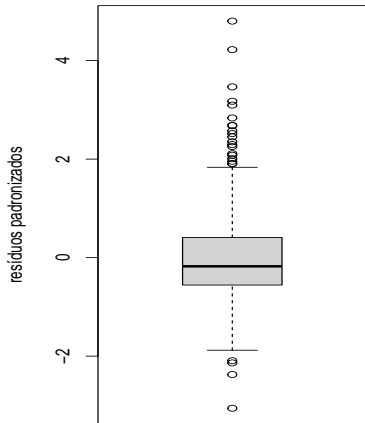
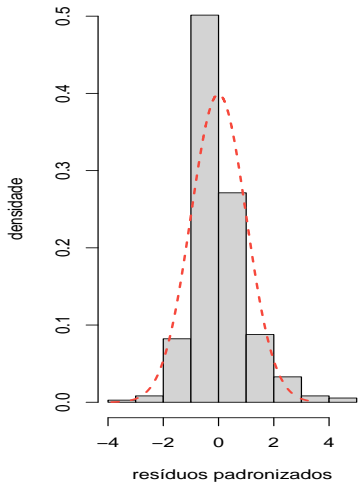


Normal Q-Q Plot of Std Residuals

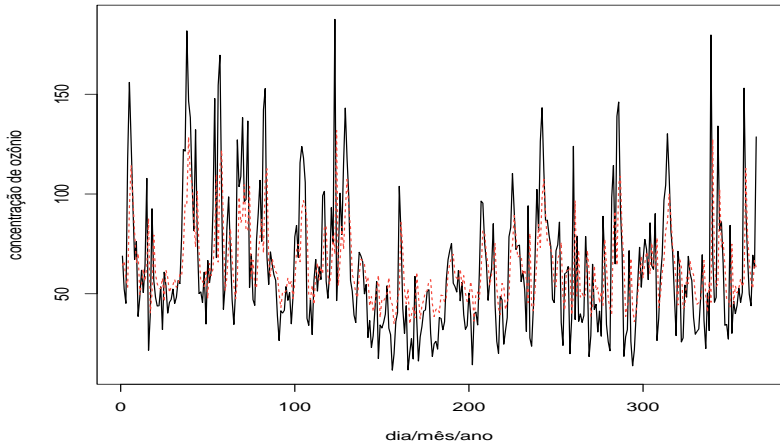


p values for Ljung-Box statistic



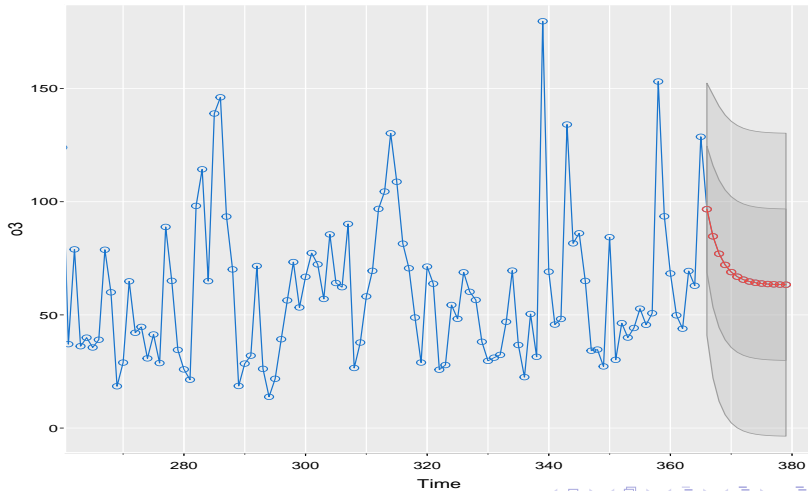


Previsão para valores observados (preto(O), vermelho (P))



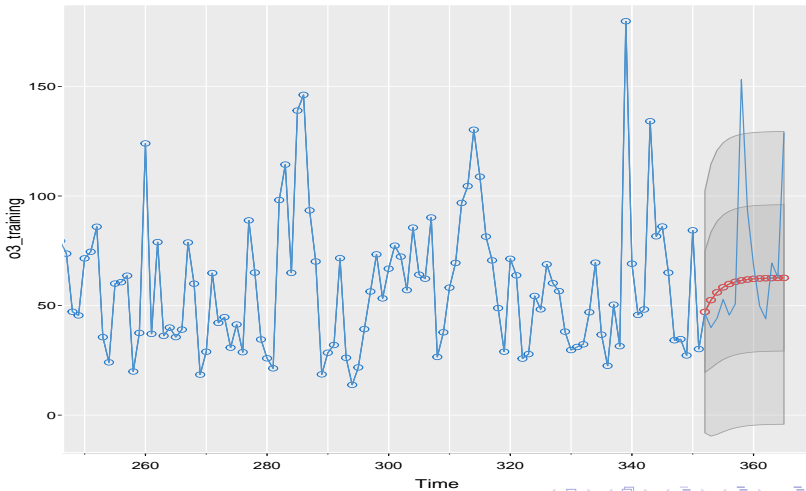
Previsão para valores futuros

previsão para duas semanas



Previsão para valores futuros observados

previsão para duas semanas



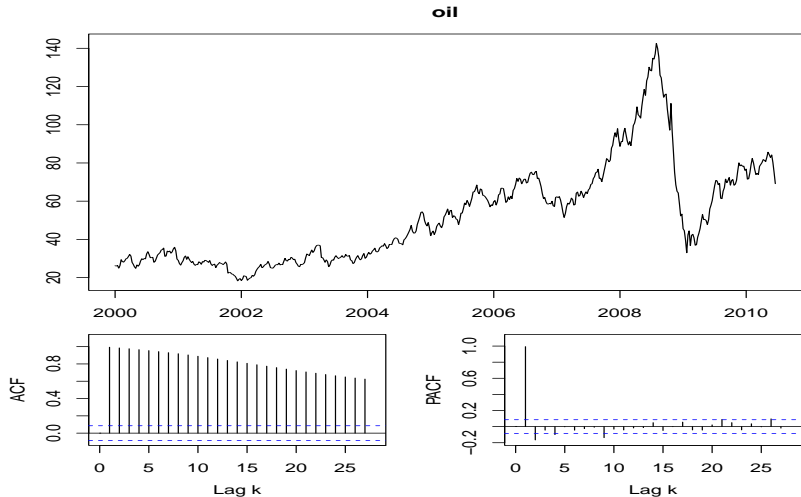
Comentários

- O modelo não se ajustou bem, do ponto de vista de análise residual, pois os resíduos apresentam assimetria à direita, com a presença de vários outliers.
- Do ponto de vista preditivo, o modelo também não se ajustou bem, pois os valores preditos (da ST observada e de valores futuros) estão bem distantes dos valores observados.
- Alternativa, ajustar um modelo com distribuição (erros/variável resposta) com assimetria à direita ou com caudas pesadas e assimetria.

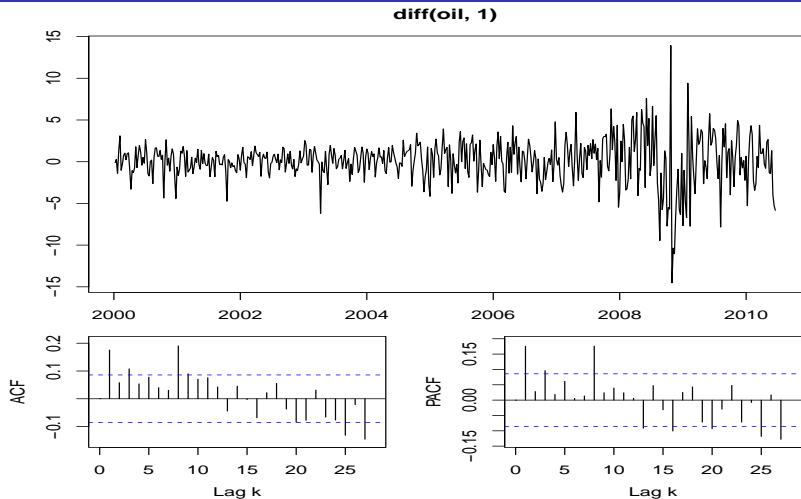
Preço (mundial) do petróleo

- Exercício: escrever os modelos ajustados.
- Petróleo bruto, preço spot WTI FOB (em dólares por barril), dados semanais de 2000 a meados de 2010 (total de $n = 545$ observações).
- ST aparentemente não estacionária. Em princípio, vamos considerar um processo $ARIMA(p,q,d)$.

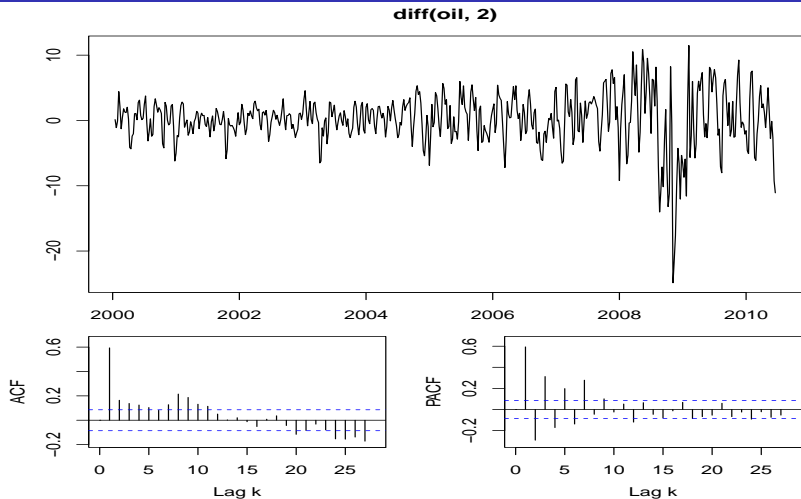
Gráficos de ST, FAC e FACP da ST



Gráficos de ST, FAC e FACP da ST (Δ_1)



Gráficos de ST, FAC e FACP da ST (Δ_2)

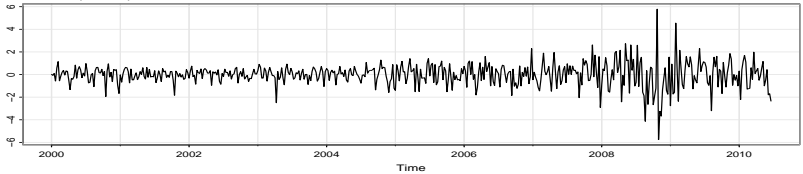


Suavização por kernel ($b=4$)

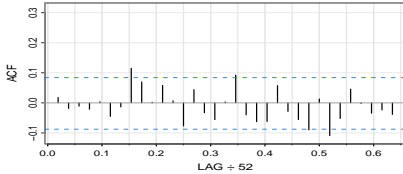


Model: (3,1,3)

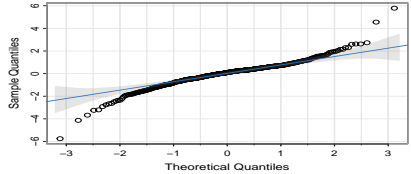
Standardized Residuals



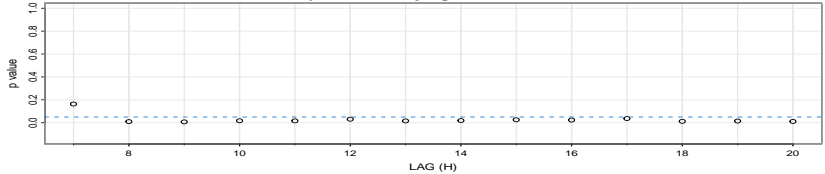
ACF of Residuals

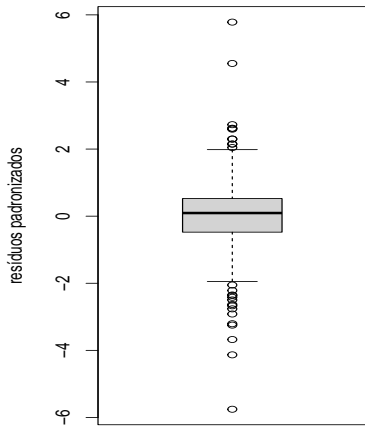
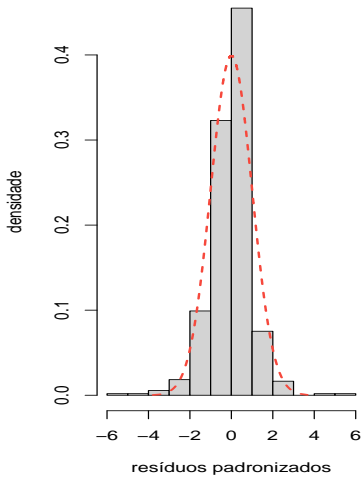


Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic



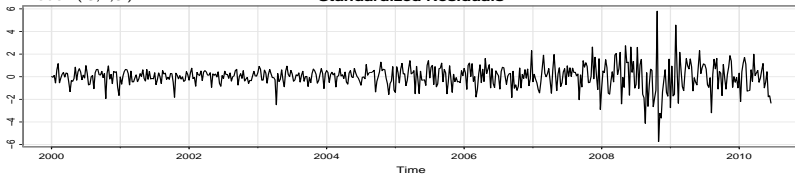


Estimativas (com intercepto)

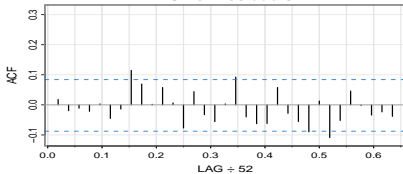
| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|------------|-------|------|-----------------|----------|---------|
| ϕ_1 | -0,88 | 0,06 | [-1,00 ; -0,76] | -14,20 | <0,0001 |
| ϕ_2 | 0,64 | 0,09 | [0,46 ; 0,82] | 7,06 | <0,0001 |
| ϕ_3 | 0,82 | 0,06 | [0,71 ; 0,93] | 14,46 | <0,0001 |
| θ_1 | 1,04 | 0,07 | [0,91 ; 1,18] | 15,31 | <0,0001 |
| θ_2 | -0,47 | 0,12 | [-0,70 ; -0,23] | -3,88 | 0,0001 |
| θ_3 | -0,78 | 0,07 | [-0,92 ; -0,65] | -11,59 | <0,0001 |
| c | 0,06 | 0,20 | [-0,33 ; 0,45] | 0,31 | 0,7603 |

Model: (3,1,3)

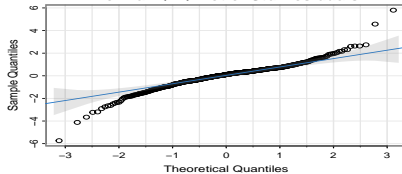
Standardized Residuals



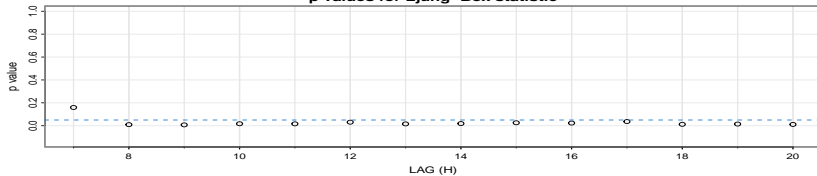
ACF of Residuals

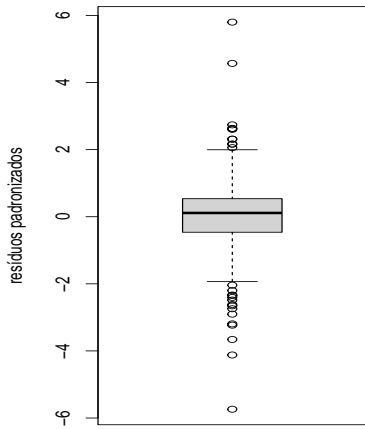
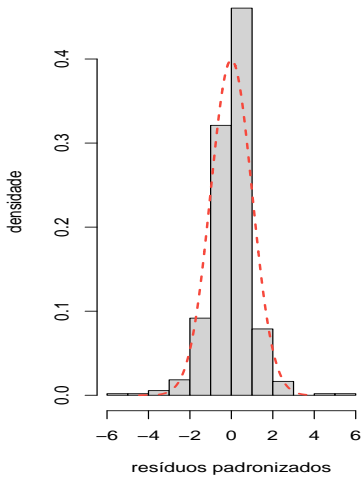


Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic



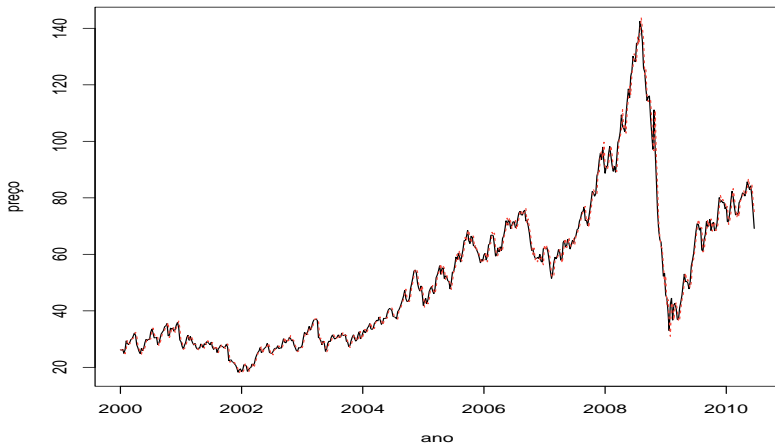


Estimativas

| Par. | Est. | EP | IC(95%) | Estat. t | p-valor |
|------------|-------|------|-----------------|----------|---------|
| ϕ_1 | -0,88 | 0,06 | [-1,00 ; -0,76] | -14,42 | <0,0001 |
| ϕ_2 | 0,64 | 0,09 | [0,47 ; 0,82] | 7,22 | <0,0001 |
| ϕ_3 | 0,82 | 0,06 | [0,71 ; 0,93] | 14,64 | <0,0001 |
| θ_1 | 1,04 | 0,07 | [0,91 ; 1,17] | 15,53 | <0,0001 |
| θ_2 | -0,47 | 0,12 | [-0,70 ; -0,24] | -3,96 | 0,0001 |
| θ_3 | -0,79 | 0,07 | [-0,92 ; -0,65] | -11,76 | <0,0001 |

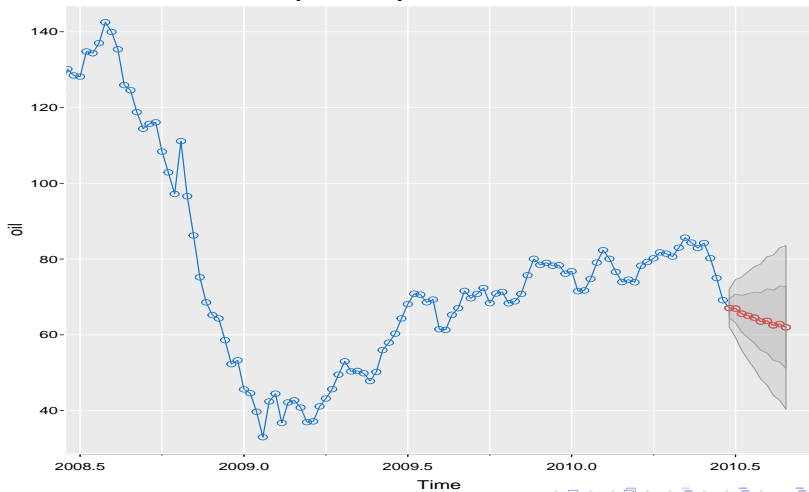
raízes: $-\phi(z) \approx (1,128; 1,042; 1,042)$ (processo estacionário e causal) e
 $\theta(z) \approx (> 1,001; > 1,001; 4,750)$ (processo invertível)

Previsão para valores observados



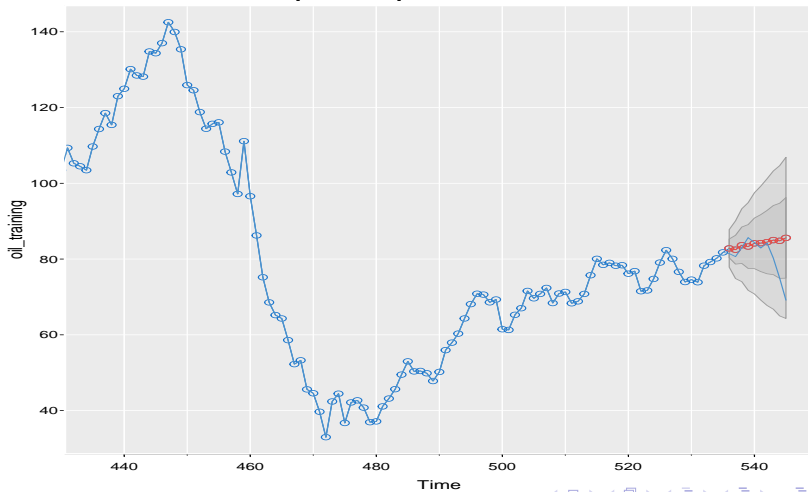
Previsão para valores futuros

previsão para 10 semanas



Previsão para valores futuros observados

previsão para 10 semanas



Comentários

- O modelo não se ajustou bem aos dados (possivelmente, pela distribuição dos erros), em relação à distribuição dos erros e o fato dos resíduos não serem ruído branco.
- Contudo, em termos de previsão dos valores observados, o modelo se comportou bem, embora para eventos futuros, novamente, observemos problemas.
- Sugestão: ajustar um modelo semelhante, mas supondo distribuição t de Student para os erros.