

Trabalho Prático 1 - Análise de Séries Temporais

Marcos Augusto Daza Barbosa, William Edward Rappel de Amorim

25 de Janeiro de 2023

1 Introdução

Neste trabalho, será realizado um estudo com o objetivo de ajustar modelos ARIMA e ETS aos dados de uma série temporal selecionada do banco de dados da competição de previsão M3, disponível no pacote `mcomp`. A série escolhida para realizar essa análise é a com `id` igual a 1931.

Ela é uma série mensal, que descreve os envios de madeira. Possui 126 observações de treino, denominadas de parte *in-sample*, e 18 observações de teste, denominadas de parte *out-of-sample*.

A figura abaixo apresenta o comportamento observado desta série na amostra de treino, que inicia em Janeiro de 1982 e termina em Junho de 1992.

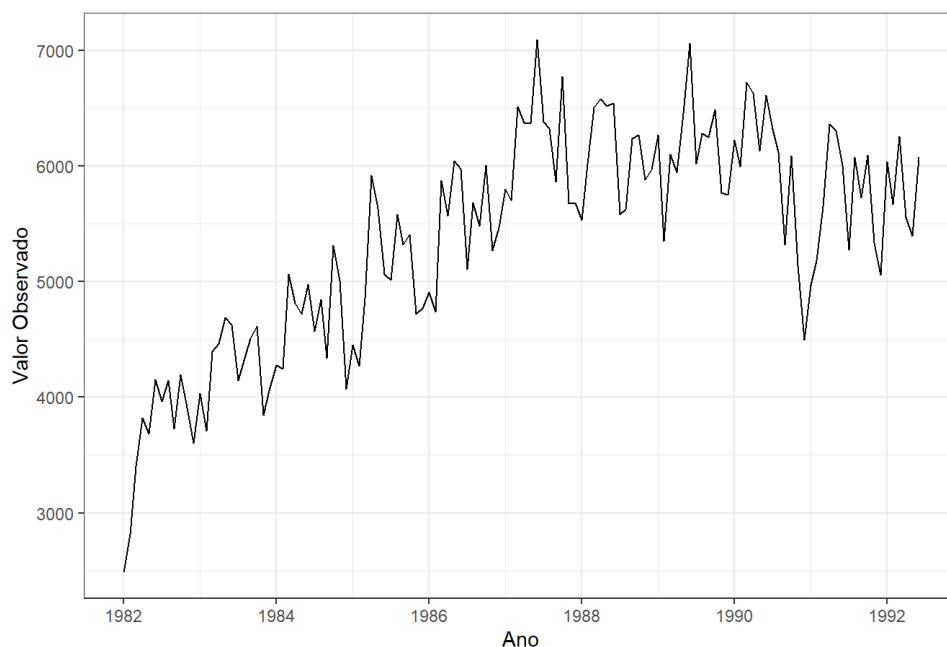


Figure 1.1: Comportamento da série ao longo do tempo.

Ao analisar a figura, parece existir uma tendência de crescimento ocorrida até 1988. Para entender mais profundamente o comportamento das componentes dessa série temporal, foi realizada uma decomposição via STL, por meio da função automática `mstl`.

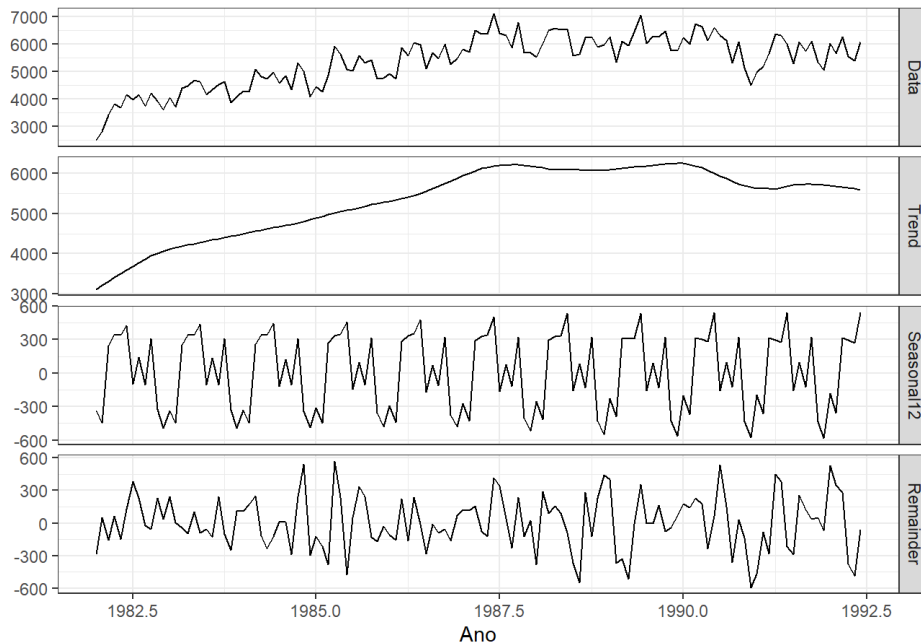


Figure 1.2: Decomposição STL da série.

A figura ilustra o resultado dessa decomposição. A componente de tendência (*Trend*) é inicialmente crescente e depois estabiliza, e é bastante relevante no valor da série (*Data*), pois sua escala varia de 3000 a 6000, enquanto a série original varia de 2000 a 7000. A componente de sazonalidade (*Seasonal12*) identificada pela função apresenta ciclo sazonal de tamanho 12, o que já era esperado por se tratar de uma série mensal. Essa componente aparenta ser aproximadamente constante ao longo da série, tendo uma escala variando de -600 a 600. Já a componente de ruído (*Remainder*) apresentou um bom comportamento na maior parte do tempo, similar a um ruído branco.

2 Modelos ARIMA

Nesta seleção, será realizado um procedimento de seleção e análise de modelos ARIMA ajustados nesta série temporal.

2.1 Transformações

O primeiro passo para determinar o melhor modelo a ser utilizado é entender quais transformações devem ser aplicadas para tornar a série original estacionária. Ao aplicar a função `ndiffs`, obtém-se o valor 1, ou seja, deve-se ser aplicada uma diferença simples e o d é igual a 1.

Em seguida, aplica-se a função `nsdiffs` ao resultado da primeira diferença simples. Ela retorna 0, ou seja, não serão necessárias diferenças sazonais. Dessa forma, D igual a 0.

A figura abaixo apresenta o comportamento observado da série original e da série diferenciada.

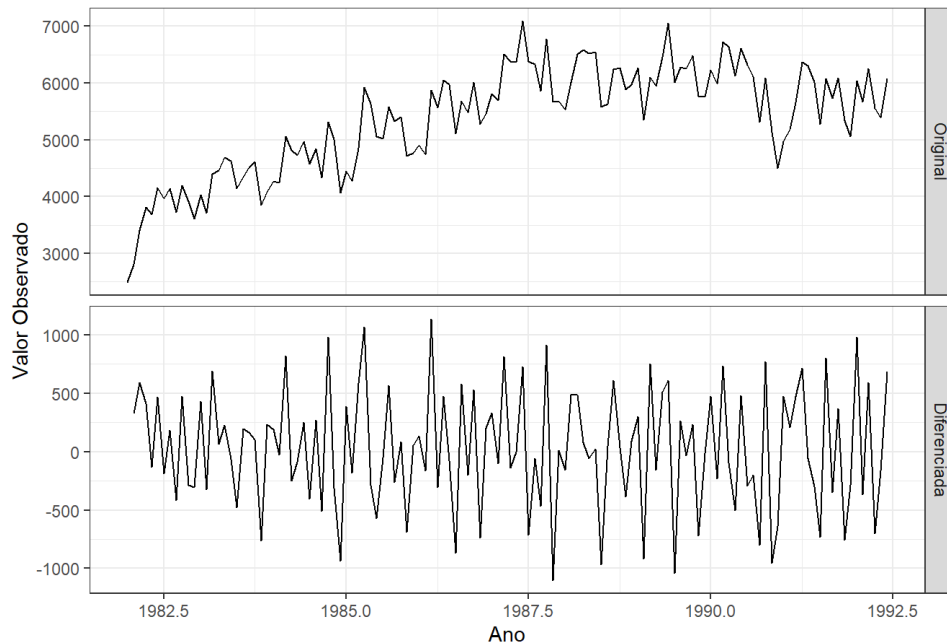


Figure 2.1: Differenciação.

A série original é nitidamente não estacionária, e a série diferenciada aparenta ser, hipótese confirmada ao realizar o teste *Augmented Dickey-Fuller*.

```
##
## Augmented Dickey-Fuller Test
##
## data: .
## Dickey-Fuller = -6.7691, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

2.2 Seleção da ordem

Para selecionar os parâmetros p , q , P e Q , deve-se investigar as autocorrelações e autocorrelações parciais.

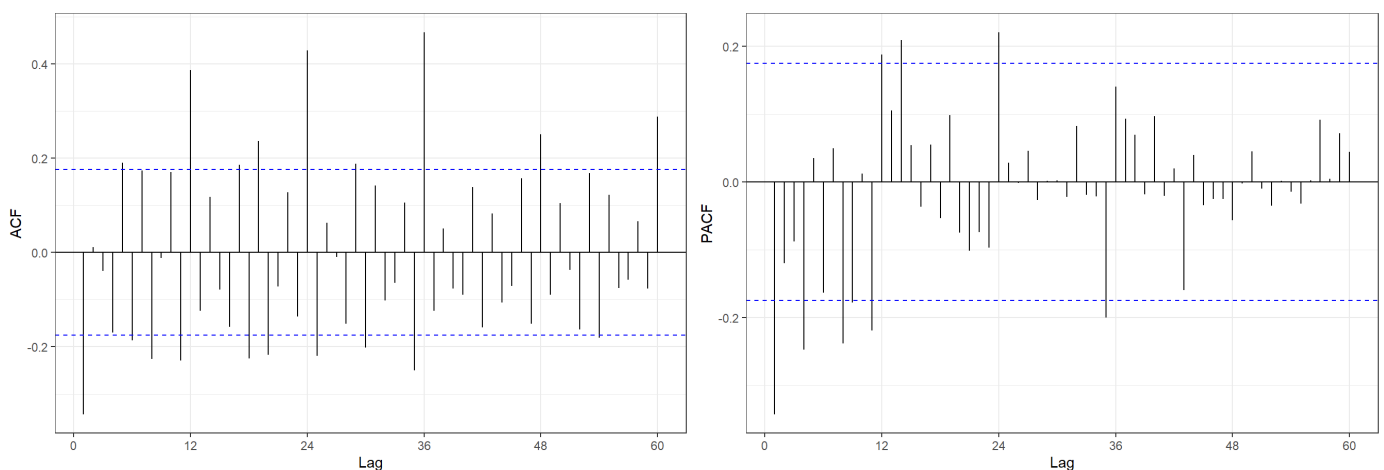


Figure 2.2: ACF e PACF.

Ao analisar o gráfico ACF, dentre as não sazonais, parece haver um valor mais significativo no primeiro *lag*, assim q igual a 1 é um bom palpite. Já ao investigar os *lags* sazonais, os mais significativos são 12, 24 e 36. Assim, serão consideradas como possibilidades para Q os valores 1, 2 e 3.

Já ao analisar o PACF, dentre os não sazonais, novamente o primeiro é o mais significativo, assim será utilizado p igual a 1. Para os sazonais, o *lag* 24 é o mais significativo. Dessa forma, serão consideradas como possibilidades para P : 0, 1 e 2.

Ao ajustar todos os modelos ARIMA correspondentes as combinações dos parâmetros p , q , P e Q , seleciona-se o modelo com menor AICc, que é apresentado abaixo.

```
## Series: .
## ARIMA(1,1,1)(1,0,2)[12]
##
## Coefficients:
##          ar1          ma1          sar1          sma1          sma2
##      0.2325   -0.7083   0.9747   -1.0288   0.2724
## s.e.  0.1337   0.0953   0.0244   0.1158   0.1025
##
## sigma^2 estimated as 112792:  log likelihood=-911.02
## AIC=1834.05   AICc=1834.76   BIC=1851.02
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -29.76627 327.7507 257.4193 -0.6064746 4.932637 0.4998602
##              ACF1
## Training set 0.006708938
```

Dessa forma, foi selecionado o modelo $ARIMA(1,1,1)(1,0,2)[12]$, com AICc de 1834,76.

Em seguida, realiza-se o mesmo procedimento de seleção pelo AICc, só que agora considerando a transformação de Box-Cox, sendo que o parâmetro λ é definido automaticamente por meio da heurística já implementada na função `Arima` do pacote `forecast`.

```
## Series: .
## ARIMA(1,1,1)(1,0,2)[12]
## Box Cox transformation: lambda= 0.6390879
##
## Coefficients:
##          ar1          ma1          sar1          sma1          sma2
##      0.2384   -0.6897   0.9768   -1.0504   0.2882
## s.e.  0.1400   0.1037   0.0213   0.1145   0.1026
##
## sigma^2 estimated as 227.9:  log likelihood=-523.72
## AIC=1059.44   AICc=1060.15   BIC=1076.41
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -35.99973 326.5198 260.4114 -0.7076345 4.977515 0.5056703
##              ACF1
## Training set -0.008303924
```

Dessa forma, foi selecionado o modelo $ARIMA(1,1,1)(1,0,2)[12]$, com AICc de 1060,15 e o λ escolhido igual a 0,64.

2.3 Análise de resíduos

Para analisar os resíduos, serão utilizadas abordagens gráficas e testes de hipóteses.

Primeiro, serão realizadas análises com o modelo ARIMA sem a transformação de Box-Cox.

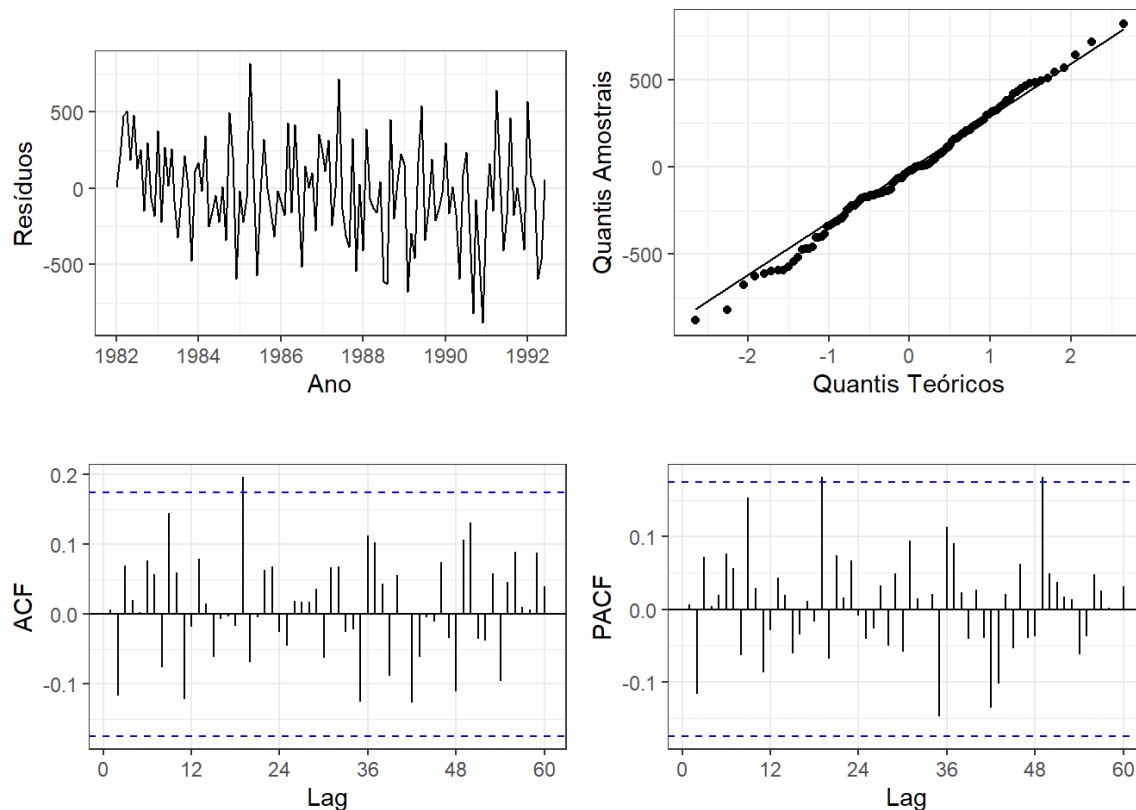


Figure 2.3: Análise de resíduos modelo ARIMA.

Ao avaliar o primeiro gráfico, os resíduos aparentam ser estacionários, tendo média zero e variância constante. Além disso, ao avaliar o *qqplot*, aparentam ser aproximadamente normais. Por último, por meio do ACF e PACF, nenhuma autocorrelação é muito significativa, sendo quase todas dentro da banda de não significância e, as poucas que saíram, estão bem próximas do limite.

Ao realizar os testes *Augmented Dickey-Fuller*, *Ljung-Box* e *Shapiro-Wilk*, conclui-se que os resíduos são estacionários, independentes e normais, respectivamente.

Agora, realiza-se procedimento semelhante, porém com o modelo ARIMA com a transformação de Box-Cox.

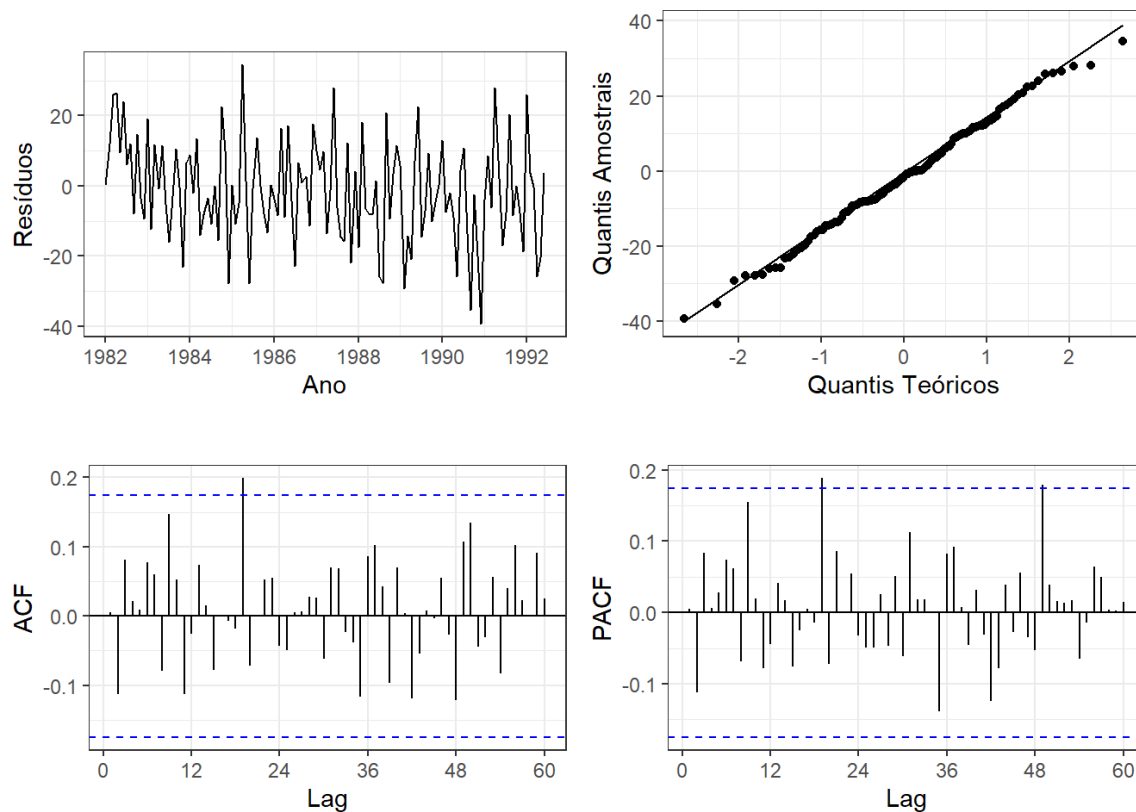


Figure 2.4: Análise de resíduos modelo ARIMA com Box-Cox.

As conclusões são muito semelhantes às obtidas anteriormente: os resíduos aparentam ser estacionários, normais e independentes. Ao realizar os testes *Augmented Dickey-Fuller*, *Ljung-Box* e *Shapiro-Wilk*, estas hipóteses são confirmadas.

3 Modelos ETS

Nesta seção, será realizado um procedimento de seleção e análise de modelos ETS ajustados nesta série temporal.

3.1 Seleção

Os modelos serão selecionados de forma automática, por meio da função `ets` do pacote `forecast`. Esse procedimento será realizado sem a transformação Box-Cox e, em seguida, com a aplicação dela considerando o parâmetro `lambda` definido automaticamente por meio da heurística já implementada na função.

```
## ETS(M,Ad,A)
##
## Call:
## ets(y = .)
##
## Smoothing parameters:
##   alpha = 0.3284
##   beta  = 1e-04
##   gamma = 1e-04
##   phi   = 0.9764
##
## Initial states:
##   l = 3247.79
##   b = 74.4494
##   s = -537.9187 -342.2722 300.1228 -128.3485 121.0667 -113.7734
##       476.2259 368.1361 361.0313 269.9528 -432.5925 -341.6303
##
##   sigma: 0.0617
##
##      AIC      AICc      BIC
## 2085.694 2092.086 2136.747
##
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
## Training set -15.52275 305.1435 242.596 -0.552088 4.632971 0.4710763 0.1504657
```

O modelo sem Box-Cox selecionado é o ETS(M, Ad, A), ou seja, com erros multiplicativos, tendência aditiva com *damped* e sazonalidade aditiva. Ele apresenta AICc de 2092,09.

```
## ETS(A,Ad,A)
##
## Call:
## ets(y = ., lambda = "auto")
##
## Box-Cox transformation: lambda= 0.6391
##
## Smoothing parameters:
##   alpha = 0.3428
##   beta  = 1e-04
##   gamma = 1e-04
##   phi   = 0.9735
##
## Initial states:
##   l = 273.3698
##   b = 3.6225
##   s = -26.2316 -17.1407 14.1811 -4.6038 5.5424 -4.8251
##       20.9924 16.4844 15.7625 12.2844 -19.345 -13.1011
##
## sigma: 14.7497
##
##      AIC      AICc      BIC
## 1305.299 1311.691 1356.352
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -7.975853 304.3027 240.8247 -0.4050985 4.597695 0.4676367
##
##              ACF1
## Training set 0.1522253
```

O modelo com Box-Cox selecionado é o ETS(A, Ad, A), ou seja, com erros aditivos, tendência aditiva com *damped* e sazonalidade aditiva. Ele apresenta AICc de 1311,69 e teve *lambda* escolhido de 0,64.

3.2 Análise de resíduos

Primeiro, serão realizadas análises com o modelo ETS sem a transformação de Box-Cox.

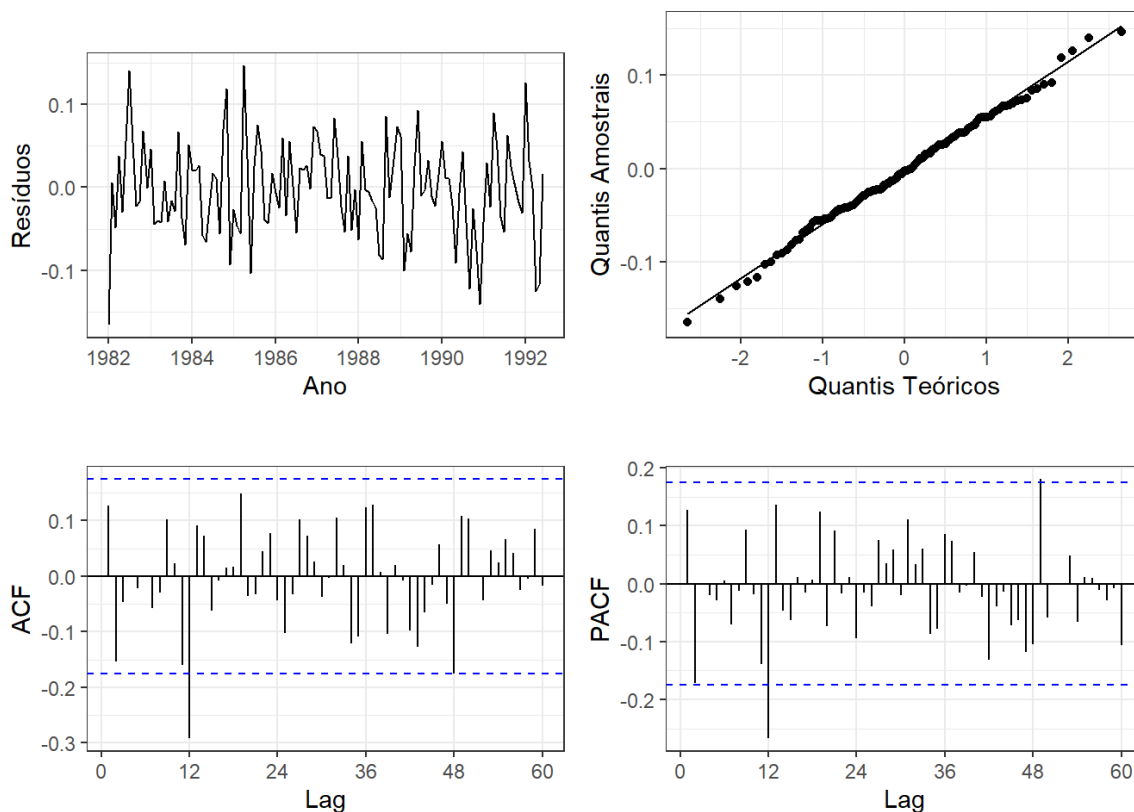


Figure 3.1: Análise de resíduos modelo ETS.

Por meio do primeiro gráfico, os resíduos aparentam ser estacionários, tendo média zero e variância constante. Já o *qqplot* demonstra provável normalidade. Porém, os gráficos ACF e PACF, apresentam um valor bastante significativo no *lag* 12. Por meio dos testes, confirma-se essas hipóteses: os resíduos são estacionários, normais e **não** independentes.

Agora, será avaliado o modelo ETS com transformação Box-Cox.

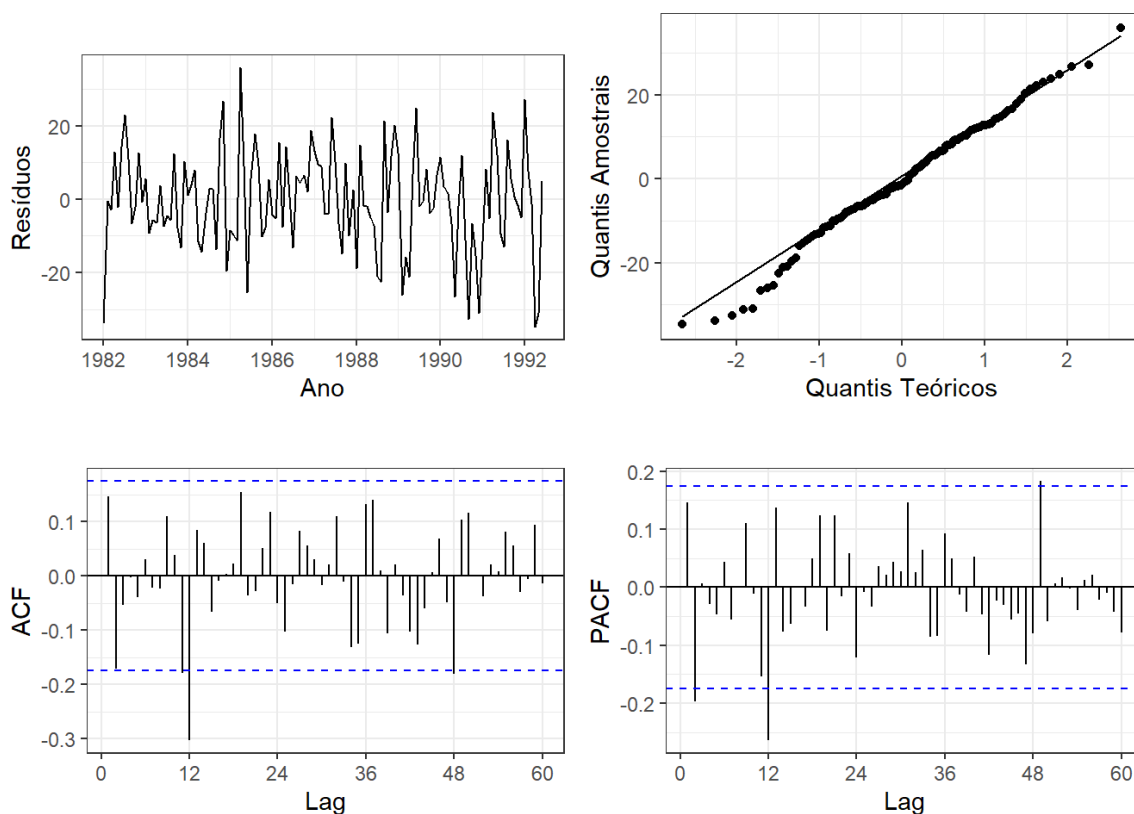


Figure 3.2: Análise de resíduos modelo ETS.

As conclusões são idênticas às obtidas com o modelo ETS sem a transformação Box-Cox: apesar de serem estacionários e normais, os resíduos apresentam **não** independência.

4 Estudo do desempenho preditivo

Nesta seção, será realizado um estudo da performance preditiva dos 4 modelos candidatos apresentados anteriormente, utilizando uma janela deslizante, iniciando no ponto $n-14$ e considerando horizontes de predição até 5 passos à frente.

A tabela e figura abaixo apresentam os resultados de erro absoluto médio obtido por cada modelo, para cada horizonte de predição.

Table 4.1: MAE por horizonte de predição.

	ARIMA	ARIMA + Box-Cox	ETS	ETS + Box-Cox
h=1	295	302	298	338
h=2	348	359	356	403
h=3	355	373	371	434
h=4	328	352	335	406
h=5	323	349	311	339

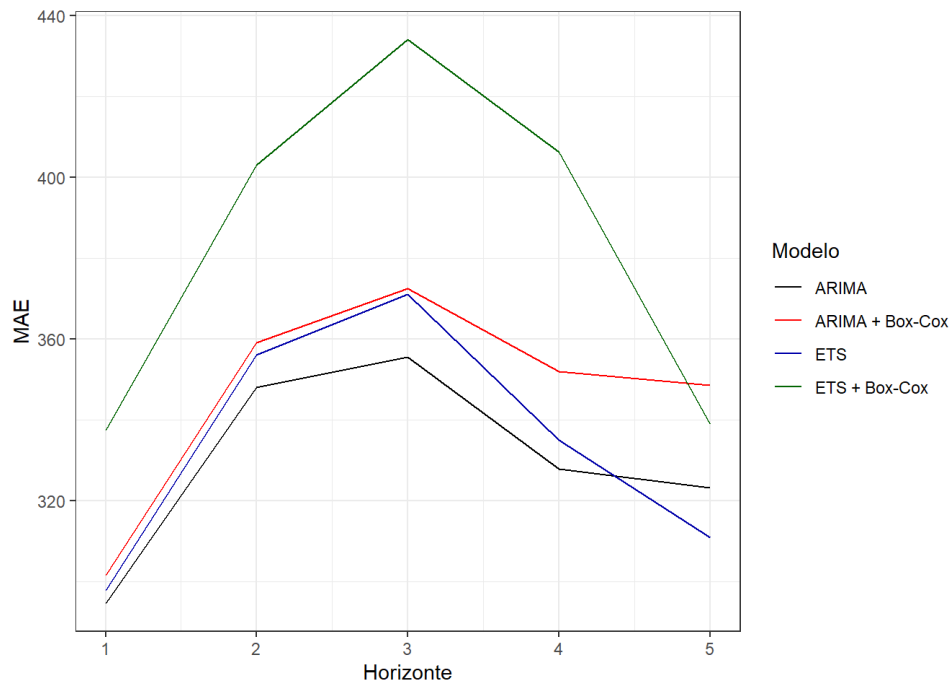


Figure 4.1: Desempenho preditivo por horizonte de predição.

A partir da análise da tabela e do gráfico, tem-se que para os primeiros 4 horizontes de predição, o modelo mais preciso é o ARIMA, seguido pelo ETS, depois pelo ARIMA + Box-Cox e, por último, o ETS + Box-Cox. Já no horizonte 5, a ordem se altera: o ETS troca com o ARIMA e o ETS + Box-Cox troca com o ARIMA + Box-Cox.

5 Resultados

Nesta seção, serão obtidas previsões pontuais e intervalares de 95% com os 4 modelos candidatos selecionados. Em seguida, o desempenho nos dados de testes serão comparados com de alguns *benchmarks*.

A tabela abaixo apresenta as previsões pontuais de cada um dos 4 modelos, além dos dados observados de teste.

Table 5.1: Previsões pontuais por horizonte de predição.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Observado	5926	5936	5568	6216	5624	5148	5328	5524	5584	5162	4880	5452	5760	5970	5426	6256	5450	5424
ARIMA	5479	5553	5404	5809	5144	5028	5276	5154	5757	5909	5887	6059	5437	5659	5411	5817	5127	4926
ARIMA + Box-Cox	5478	5544	5404	5804	5152	5041	5270	5154	5753	5913	5900	6061	5444	5665	5421	5825	5142	4939
ETS	5424	5662	5416	5848	5208	5016	5215	5127	5832	5926	5936	6047	5459	5697	5450	5881	5241	5047
ETS + Box-Cox	5415	5650	5424	5851	5152	4957	5245	5111	5819	5900	5919	6025	5440	5675	5449	5875	5175	4979

Já a tabela abaixo apresenta as previsões intervalares com 95% de confiança.

Table 5.2: Previsões intervalares de 95% de confiança por horizonte de predição.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Observado	5926	5936	5568	6216	5624	5148	5328	5524	5584	5162	4880	5452	5760	5970	5426	6256	5450	5424
ARIMA Inf	4821	4810	4612	4977	4275	4124	4338	4183	4754	4875	4824	4966	4323	4522	4250	4632	3918	3694
ARIMA Sup	6138	6296	6195	6641	6013	5933	6214	6125	6759	6942	6950	7151	6550	6796	6572	7003	6335	6157
ARIMA + Box-Cox Inf	4831	4805	4621	4956	4303	4163	4342	4200	4724	4841	4797	4915	4326	4507	4259	4606	3958	3753

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ARIMA + Box-Cox Sup	6154	6320	6230	6700	6055	5979	6262	6176	6852	7061	7083	7290	6652	6917	6682	7144	6433	6238
ETS Inf	4768	4944	4690	5045	4437	4235	4391	4284	4902	4959	4939	5013	4445	4641	4388	4764	4146	3946
ETS Sup	6080	6381	6142	6651	5980	5797	6040	5970	6763	6894	6933	7081	6474	6753	6512	6997	6335	6149
ETS + Box-Cox Inf	4785	4974	4726	5100	4407	4194	4438	4285	4925	4976	4968	5043	4472	4668	4436	4810	4140	3940
ETS + Box-Cox Sup	6072	6356	6157	6639	5939	5765	6100	5987	6765	6880	6929	7069	6475	6751	6535	7015	6290	6103

As figuras abaixo ilustram o comportamento dessas previsões, conjuntamente com o comportamento observado nos dados de teste.

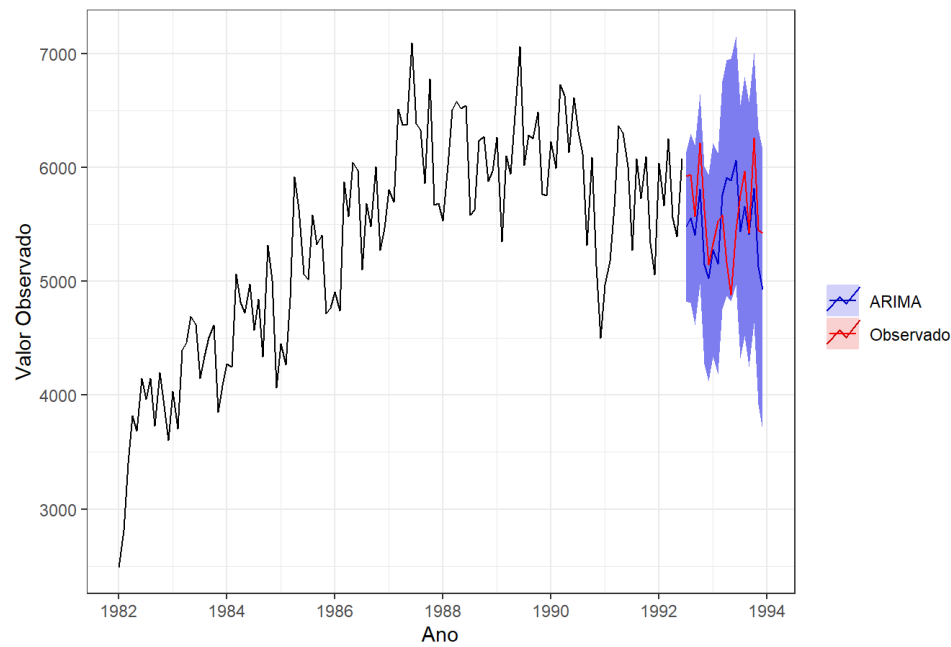


Figure 5.1: Previsão pontual e intervalar do ARIMA.

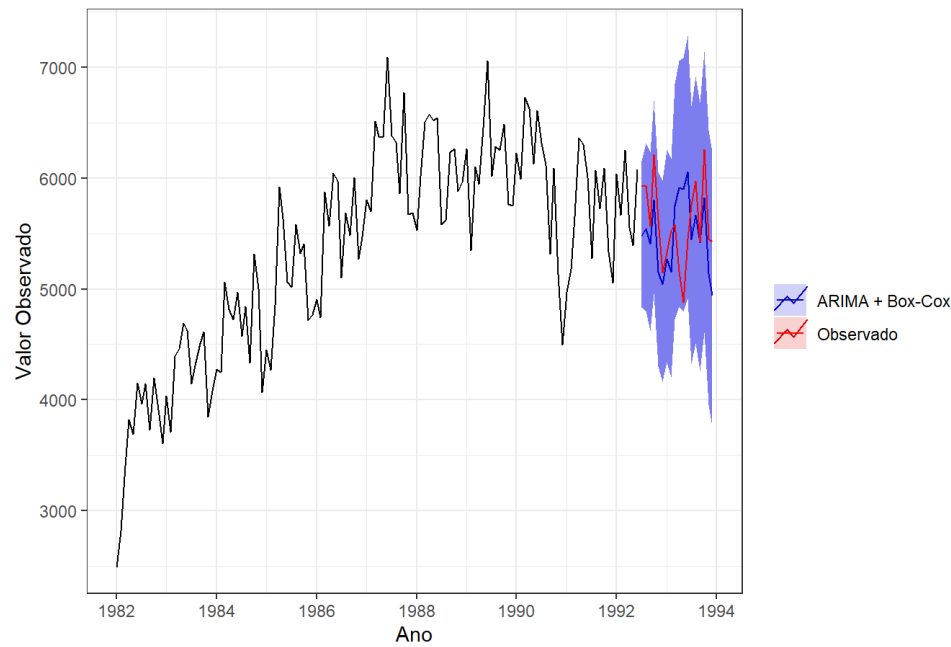


Figure 5.2: Previsão pontual e intervalar do ARIMA + Box-Cox.

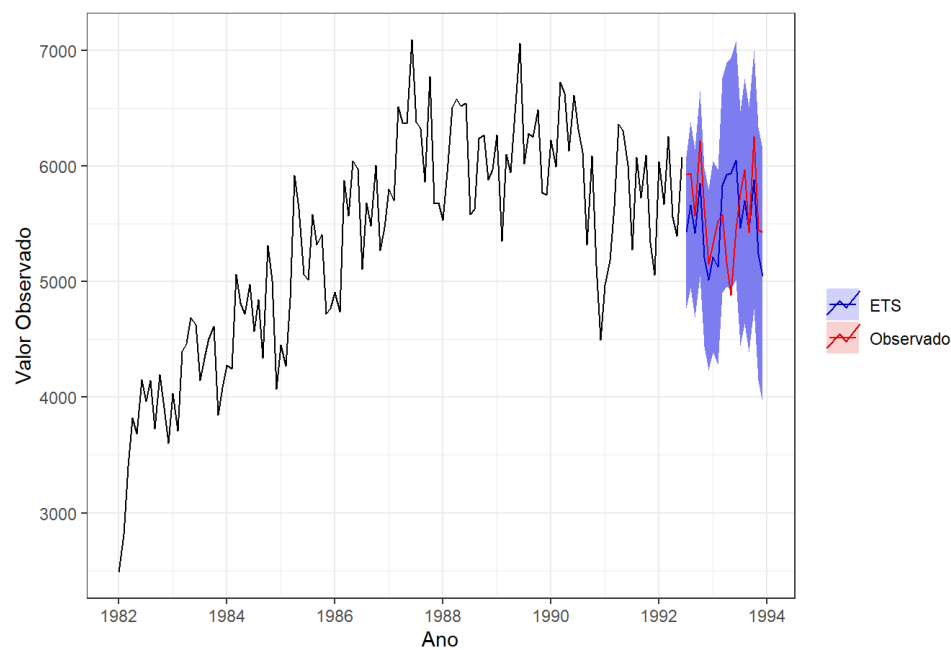


Figure 5.3: Previsão pontual e intervalar do ETS.

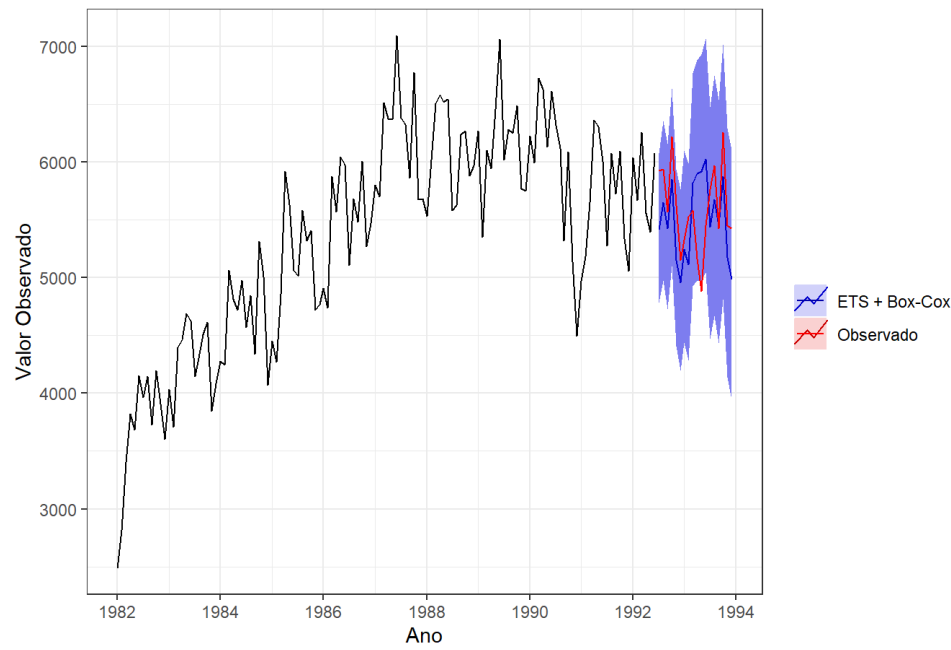


Figure 5.4: Previsão pontual e intervalar do ETS + Box-Cox.

Por meio da análise dos gráficos, todos os modelos apresentaram previsões pontuais razoáveis e previsões intervalares com bom grau de cobertura dos dados verdadeiros de teste.

Agora, serão realizadas as comparações com os *benchmarks*.

Table 5.3: MAE nos dados de teste.

	MAE
ARIMA	381
ARIMA + Box-Cox	379
ETS	365
ETS + Box-Cox	377
auto.arima	499

	MAE
SES	367
Holt	518
stlf	385
BATS	332
TBATS	332
Bagged ETS	313

Ao avaliar a tabela acima, percebe-se que apesar do ARIMA ter tido o melhor desempenho na análise de janela deslizante, o candidato com melhor desempenho no teste foi o ETS. Além disso, o ARIMA + Box-Cox foi mais preciso do que o ARIMA sem a transformação. O ETS foi superior a 4 *benchmarks*: `auto.arima`, `ses`, `holt` e `stlf`. Já BATS, TBATS e *Bagged* ETS foram superiores ao ETS, sendo que o *Bagged* ETS foi consideravelmente mais preciso que todos os demais.

6 Conclusão

A partir dos 4 modelos candidatos obtidos, os 2 modelos ARIMA apresentaram resíduos com todas as características esperadas, já os resíduos dos 2 modelos ETS apresentaram não independência.

Em relação à análise do desempenho preditivo obtido pela janela deslizante, em geral os modelos com transformação Box-Cox apresentaram pior performance do que os sem essa transformação. Além disso, os ARIMAs foram mais precisos do que os ETS.

Contudo, ao avaliar o desempenho preditivo por meio dos valores de teste, os modelos ETS foram mais precisos que ambos os ARIMAs, sendo que o ETS sem a transformação Box-Cox teve o melhor desempenho dentre os 4 candidatos.

Já na comparação com os *benchmarks*, o BATS, TBATS e *Bagged* ETS foram superiores aos 4 modelos candidatos e aos demais *benchmarks*, com destaque para o *Bagged* ETS, que teve o menor MAE dentre todos.

7 Anexo: Código R

```
### Reproducibility
set.seed(10)

### Packages
pacman::p_load(Mcomp, tidyverse, forecast, tseries, knitr, cowplot, kableExtra, forecTheta)

### Data
data(M3)
id <- 1931
period <- M3[[id]]$period
description <- M3[[id]]$description
x <- M3[[id]]$x
n <- M3[[id]]$n
h <- M3[[id]]$h
xx <- M3[[id]]$xx

### Time series plot
x %>%
  autoplot() +
  labs(x='Ano', y='Valor Observado') +
  theme_bw()

### STL decomposition
x %>%
  mstl() %>%
  autoplot() +
  labs(x='Ano') +
  theme_bw()

### ARIMA
# Diffs
diffs <- c()
if (x %>% ndiffs() == 1) {
  w <- diff(x)
  diffs <- c(diffs, 1)
  if (w %>% nsdiffs() == 1) {
    w <- diff(w, lag=12)
    diffs <- c(diffs, 12)
  }
}
cbind(Original=x, Diferenciada=w) %>%
  autoplot(facets=T) +
  labs(x='Ano', y='Valor Observado') +
  theme_bw()
w %>% adf.test()
# Manual selection
w %>%
  ggAcf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
w %>%
  ggPacf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
# p = 1
# q = 1
# P = 0, 1 ou 2
# Q = 1, 2 ou 3
p <- 1
```

```

q <- 1
P <- 0:2
Q <- 1:3
# Without Box-Cox
best_aicc <- Inf
for (Pi in P) {
  for (Qi in Q) {
    mod <- x %>% Arima(order=c(1, 1, 1), seasonal=c(Pi, 0, Qi), include.mean=F, lambda=NULL)
    if (mod$aicc < best_aicc) {
      best_mod <- mod
      best_aicc <- mod$aicc
    }
  }
}
mod1 <- best_mod
mod1 %>% summary()
# With Box-Cox
best_aicc <- Inf
for (Pi in P) {
  for (Qi in Q) {
    mod <- x %>% Arima(order=c(1, 1, 1), seasonal=c(Pi, 0, Qi), include.mean=F, lambda='auto')
    if (mod$aicc < best_aicc) {
      best_mod <- mod
      best_aicc <- mod$aicc
    }
  }
}
mod2 <- best_mod
mod2 %>% summary()
# Residual analysis without Box-Cox
res1 <- mod1 %>% residuals()
est1 <- res1 %>%
  autoplot() +
  labs(x='Ano', y='Resíduos') +
  theme_bw()
qq1 <- res1 %>%
  data.frame() %>%
  ggplot(aes(sample=res1)) +
  stat_qq() +
  stat_qq_line() +
  labs(x='Quantis Teóricos', y='Quantis Amostrais') +
  theme_bw()
acf1 <- res1 %>%
  ggAcf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
pacf1 <- res1 %>%
  ggPacf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
plot_grid(est1, qq1, acf1, pacf1, nrow=2)
res1 %>% adf.test()
res1 %>% Box.test(lag=20, type='Ljung-Box', fitdf=5)
res1 %>% shapiro.test()
mod1 %>% checkresiduals()
# Residual analysis with Box-Cox
res2 <- mod2 %>% residuals()
est2 <- res2 %>%
  autoplot() +
  labs(x='Ano', y='Resíduos') +
  theme_bw()
qq2 <- res2 %>%
  data.frame() %>%
  ggplot(aes(sample=res2)) +
  stat_qq() +

```

```

stat_qq_line() +
labs(x='Quantis Teóricos', y='Quantis Amostrais') +
theme_bw()
acf2 <- res2 %>%
  ggAcf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
pacf2 <- res2 %>%
  ggPacf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
plot_grid(est2, qq2, acf2, pacf2, nrow=2)
res2 %>% adf.test()
res2 %>% Box.test(lag=20, type='Ljung-Box', fitdf=5)
res2 %>% shapiro.test()
mod2 %>% checkresiduals()

```

```

### ETS

```

```

# Auto selection without Box-Cox

```

```

mod3 <- x %>% ets()

```

```

mod3 %>% summary()

```

```

# Auto selection with Box-Cox

```

```

mod4 <- x %>% ets(lambda='auto')

```

```

mod4 %>% summary()

```

```

# Residual analysis without Box-Cox

```

```

res3 <- mod3 %>% residuals()

```

```

est3 <- res3 %>%

```

```

  autoplot() +

```

```

  labs(x='Ano', y='Resíduos') +

```

```

  theme_bw()

```

```

qq3 <- res3 %>%

```

```

  data.frame() %>%

```

```

  ggplot(aes(sample=res3)) +

```

```

  stat_qq() +

```

```

  stat_qq_line() +

```

```

  labs(x='Quantis Teóricos', y='Quantis Amostrais') +

```

```

  theme_bw()

```

```

acf3 <- res3 %>%

```

```

  ggAcf(lag.max=12*5) +

```

```

  labs(title='') +

```

```

  theme_bw()

```

```

pacf3 <- res3 %>%

```

```

  ggPacf(lag.max=12*5) +

```

```

  labs(title='') +

```

```

  theme_bw()

```

```

plot_grid(est3, qq3, acf3, pacf3, nrow=2)

```

```

res3 %>% adf.test()

```

```

res3 %>% Box.test(lag=20, type='Ljung-Box', fitdf=17)

```

```

res3 %>% shapiro.test()

```

```

mod3 %>% checkresiduals()

```

```

# Residual analysis with Box-Cox

```

```

res4 <- mod4 %>% residuals()

```

```

est4 <- res4 %>%

```

```

  autoplot() +

```

```

  labs(x='Ano', y='Resíduos') +

```

```

  theme_bw()

```

```

qq4 <- res4 %>%

```

```

  data.frame() %>%

```

```

  ggplot(aes(sample=res4)) +

```

```

  stat_qq() +

```

```

  stat_qq_line() +

```

```

  labs(x='Quantis Teóricos', y='Quantis Amostrais') +

```

```

  theme_bw()

```

```

acf4 <- res4 %>%

```

```

  ggAcf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
pacf4 <- res4 %>%
  ggPacf(lag.max=12*5) +
  labs(title='') +
  theme_bw()
plot_grid(est4, qq4, acf4, pacf4, nrow=2)
res4 %>% adf.test()
res4 %>% Box.test(lag=20, type='Ljung-Box', fitdf=17)
res4 %>% shapiro.test()
mod4 %>% checkresiduals()

### Sliding window validation
f_arima1 <- function(y, h){
  fit <- Arima(y, order=c(1, 1, 1), seasonal=c(1, 0, 2), include.mean=F, lambda=NULL)
  forecast(fit, h)
}
f_arima2 <- function(y, h){
  fit <- Arima(y, order=c(1, 1, 1), seasonal=c(1, 0, 2), include.mean=F, lambda='auto')
  forecast(fit, h)
}
f_ets1 <- function(y, h){
  fit <- ets(y)
  forecast(fit, h)
}
f_ets2 <- function(y, h){
  fit <- ets(y, lambda='auto')
  forecast(fit, h)
}
CV_arima1 <- x %>% tsCV(forecastfunction=f_arima1, h=5, initial=n-14)
CV_arima2 <- x %>% tsCV(forecastfunction=f_arima2, h=5, initial=n-14)
CV_ets1 <- x %>% tsCV(forecastfunction=f_ets1, h=5, initial=n-14)
CV_ets2 <- x %>% tsCV(forecastfunction=f_ets2, h=5, initial=n-14)
MAE_arima1 <- CV_arima1 %>% abs() %>% colMeans(na.rm=T)
MAE_arima2 <- CV_arima2 %>% abs() %>% colMeans(na.rm=T)
MAE_ets1 <- CV_ets1 %>% abs() %>% colMeans(na.rm=T)
MAE_ets2 <- CV_ets2 %>% abs() %>% colMeans(na.rm=T)
tab <- cbind(MAE_arima1, MAE_arima2, MAE_ets1, MAE_ets2)
tab %>%
  kable(
    col.names=c('ARIMA', 'ARIMA + Box-Cox', 'ETS', 'ETS + Box-Cox'),
    caption='MAE por horizonte de predição.',
    digits=0,
    format.args=list(decimal.mark=',', scientific=F),
    align='c'
  ) %>%
  kable_styling(
    position='center',
    bootstrap_options=c('striped', 'hover', 'condensed', 'responsive')
  )
tab_plot <- tab %>%
  as.data.frame() %>%
  mutate(Horizonte=1:5) %>%
  gather(key='Modelo', value='MAE', -Horizonte)
tab_plot %>%
  ggplot(aes(x=Horizonte, y=MAE)) +
  geom_line(aes(color=Modelo)) +
  scale_color_manual(
    values=c('black', 'red', '#0000AA', 'darkgreen'),
    breaks=c('MAE_arima1', 'MAE_arima2', 'MAE_ets1', 'MAE_ets2'),
    labels=c('ARIMA', 'ARIMA + Box-Cox', 'ETS', 'ETS + Box-Cox')
  ) +
  theme_bw()

```



```

### Forecast
# tables
preds1 <- forecast(mod1, h=h, level=95)
preds2 <- forecast(mod2, h=h, level=95)
preds3 <- forecast(mod3, h=h, level=95)
preds4 <- forecast(mod4, h=h, level=95)
pontual <- t(cbind(xx, preds1$mean, preds2$mean, preds3$mean, preds4$mean))
colnames(pontual) <- 1:h
row.names(pontual) <- c('Observado', 'ARIMA', 'ARIMA + Box-Cox', 'ETS', 'ETS + Box-Cox')
pontual %>%
  kable(
    caption='Previsões pontuais por horizonte de predição.',
    digits=0,
    format.args=list(decimal.mark=',', scientific=F),
    align='c'
  ) %>%
  kable_styling(
    position='center',
    bootstrap_options=c('striped', 'hover', 'condensed', 'responsive')
  )
intervalares <- t(cbind(xx, preds1$lower, preds1$upper, preds2$lower, preds2$upper,
  preds3$lower, preds3$upper, preds4$lower, preds4$upper))
colnames(intervalares) <- 1:h
row.names(intervalares) <- c('Observado', 'ARIMA Inf', 'ARIMA Sup', 'ARIMA + Box-Cox Inf',
  'ARIMA + Box-Cox Sup', 'ETS Inf', 'ETS Sup', 'ETS + Box-Cox Inf',
  'ETS + Box-Cox Sup')
intervalares %>%
  kable(
    caption='Previsões intervalares de 95% de confiança por horizonte de predição.',
    digits=0,
    format.args=list(decimal.mark=',', scientific=F),
    align='c'
  ) %>%
  kable_styling(
    position='center',
    bootstrap_options=c('striped', 'hover', 'condensed', 'responsive')
  )
# plots
plot_preds <- function(mod, nome='') {
  vec <- c(nome, 'Observado')
  cores <- c('#0000AA', 'red')
  names(cores) <- vec
  preds <- forecast(mod, h=h, level=95)
  plot_obj <- x %>%
    autoplot() + xlab('Ano') + ylab('Valor Observado') + theme_bw() +
    autolayer(preds, series=nome) +
    autolayer(xx, series='Observado') +
    scale_colour_manual(
      values=cores,
      breaks=vec,
      name='')
  return(plot_obj)
}
plot_preds(mod1, 'ARIMA')
plot_preds(mod2, 'ARIMA + Box-Cox')
plot_preds(mod3, 'ETS')
plot_preds(mod4, 'ETS + Box-Cox')
# Benchmark comparison
preds <- list(
  'ARIMA' = forecast(mod1, h=h),
  'ARIMA + Box-Cox' = forecast(mod2, h=h),
  'ETS' = forecast(mod3, h=h),
  'ETS + Box-Cox' = forecast(mod4, h=h),

```

```
'auto.arima' = forecast(auto.arima(x), h=h),
'SES' = ses(x, h=h),
'Holt' = holt(x, h=h),
'sltf' = stlf(x, h=h),
'BATS' = forecast(bats(x), h=h),
'TBATS' = forecast(tbats(x), h=h),
'Bagged ETS' = forecast(baggedETS(x), h=h)
)
mae <- unlist(lapply(preds, function(m) return(mean(abs(xx - m$mean)))))
final <- data.frame(MAE=mae)
final %>%
  kable(
    caption='MAE nos dados de teste.',
    digits=0,
    format.args=list(decimal.mark=',', scientific=F),
    align='c'
  ) %>%
  kable_styling(
    position='center',
    bootstrap_options=c('striped', 'hover', 'condensed', 'responsive')
  )
```