

# Trabalho Prático 2

## Análise de Séries Temporais - 1/2023

Ana Tércia Freires da Silva -

Gabriel Vêras Monteiro - 19/0106794

Gabriela Carneiro de Almeida - 18/0120816

### Table of contents

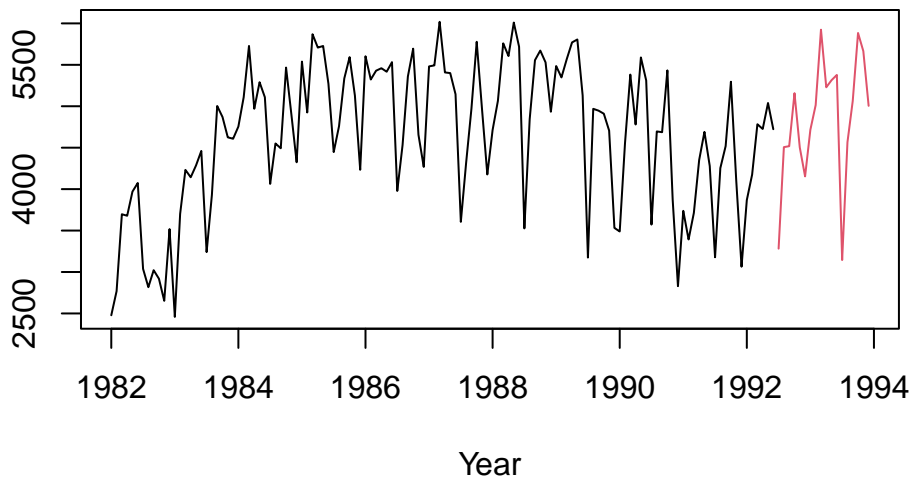
<b>Introdução: série selecionada, características e decomposição</b>	<b>2</b>
<b>Modelos ARIMA: seleção, transformações e resíduos</b>	<b>4</b>
Sem transformação de Box-Cox . . . . .	4
Com transformação de Box-Cox . . . . .	11
<b>Modelos ETS: seleção, transformações e resíduos</b>	<b>17</b>
Resíduos . . . . .	18
Modelo com transformação . . . . .	20
Seleção . . . . .	20
Resíduos . . . . .	22
<b>Estudo de desempenho preditivo</b>	<b>23</b>
Resultados da Janela Deslizante . . . . .	23
Performance em relação aos horizontes de previsão . . . . .	23
ARIMA . . . . .	23
ETS . . . . .	23
<b>Resultados</b>	<b>23</b>
<b>Apêndice</b>	<b>24</b>

## Introdução: série selecionada, características e decomposição

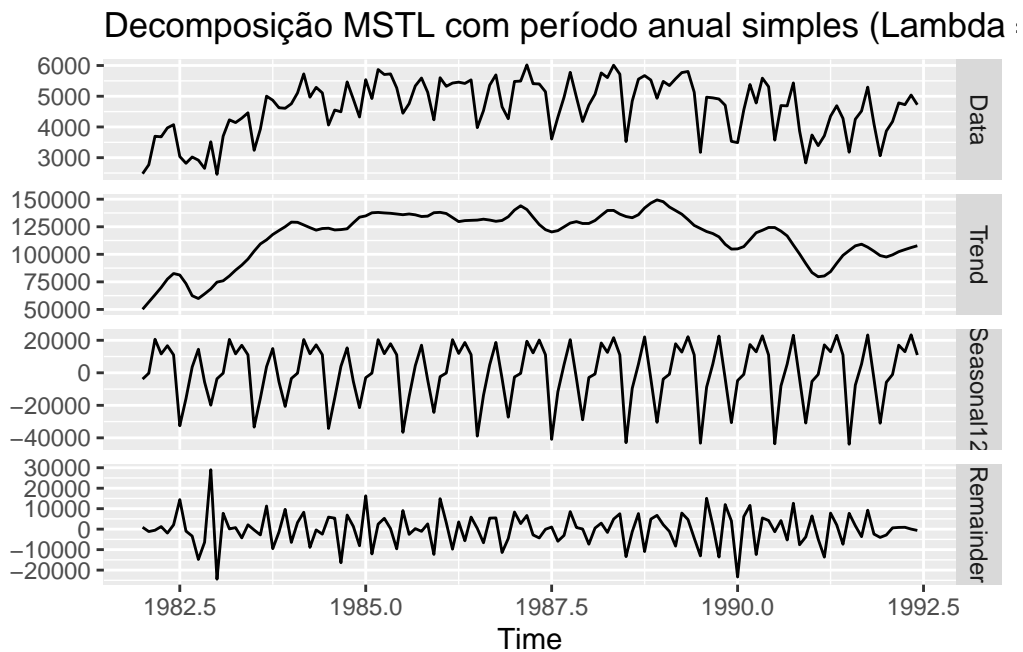
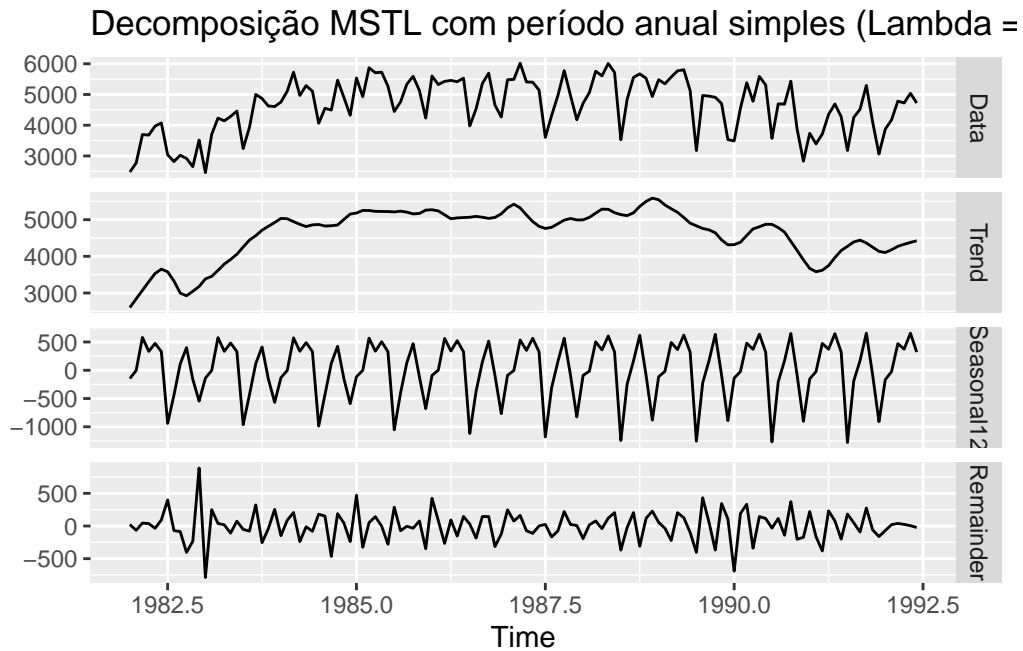
A série temporal escolhida foi a de número *id* correspondente a 1891. De acordo com a definição do próprio pacote, refere-se a *Pneumatic casings, original equipment*. Foram realizadas medidas mensais de 1982 a 1992 e o horizonte de previsão requerido é das 18 ocorrências seguintes.

O gráfico da série, com *in* e *out-sample*, é exposto a seguir.

### Série Temporal M3–1891



Após visualização inicial da série, foi feita sua decomposição via MSTL. Foram feitas duas decomposições, uma ajustando  $\lambda = \text{"auto"}$  e  $\lambda = \text{NULL}$ , resultando em duas composições muito similares.



Aplicando a decomposição MSTL com sazonalidade anual, é possível visualizar uma tendência de crescimento no início das medições, passando por uma estabilização e posterior decréscimo. Ao final, a série parece retomar um pequeno crescimento. Além disso, é possível notar que a série tem um componente sazonal mensal, com ciclos claros. Por fim, o ruído parece se

aproximar de um padrão de ruído branco.

## Modelos ARIMA: seleção, transformações e resíduos

### Sem transformação de Box-Cox

```
KPSS Test for Level Stationarity
```

```
data:  serie_ms
```

```
KPSS Level = 0.52764, Truncation lag parameter = 4, p-value = 0.03544
```

```
[1] 1
```

```
[1] 1
```

```
KPSS Test for Level Stationarity
```

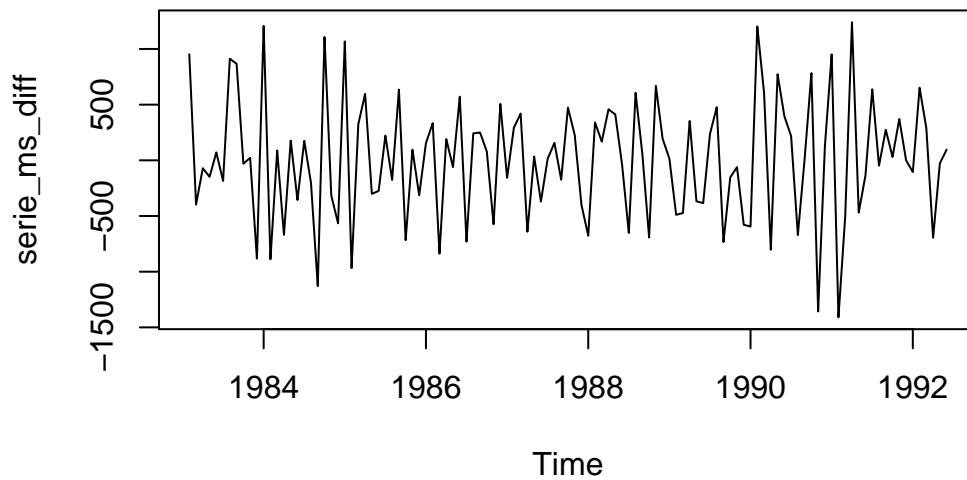
```
data:  serie_ms_diff
```

```
KPSS Level = 0.044309, Truncation lag parameter = 4, p-value = 0.1
```

Inicialmente foi feito o teste de estacionaridade KPSS na série segundo as hipóteses:

$$\begin{cases} H_0 : \text{O processo é estacionário} \\ H_1 : \text{O processo possui raiz unitária} \end{cases} \quad (1)$$

O teste indica que a série não é estacionária ( $p\text{-valor} = 0,03544$ ), portanto, é necessário aplicar diferenciações para tornar a série estacionária antes de ajustar um modelo adequado que se ajuste a série em análise. O número de diferenciações simples necessárias é igual a um, dessa forma, a estimativa para ‘ $d$ ’ do modelo SARIMA é de ‘ $d$ ’ = 1. Além disso, em se tratando de uma série sazonal, é necessário verificar o número de diferenciações sazonais para retirar o efeito da sazonalidade. Analogamente, foi obtido o valor de uma diferenciação sazonal necessária, se tratando, portanto, do ‘ $D$ ’ do modelo SARIMA, sendo ‘ $D$ ’ = 1. Para a diferenciação sazonal, pelo ciclo sazonal ser igual a 12, é necessário a utilização do lag sazonal igual a 12 ( $\text{lag} = 12$ ).

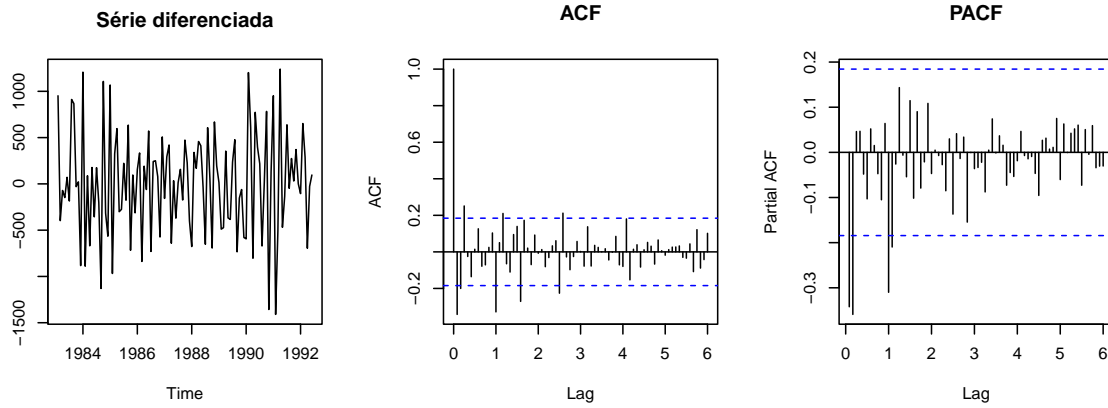


#### KPSS Test for Level Stationarity

```
data:  serie_ms_diff
KPSS Level = 0.044309, Truncation lag parameter = 4, p-value = 0.1
```

O gráfico acima mostra o comportamento da série após as diferenciações simples e sazonal. A série aparenta ter comportamento estacionário, o que foi confirmado a partir da aplicação do teste KPSS, cujas hipóteses já foram explicitadas anteriormente. Por meio do teste foi possível notar que a série realmente está estacionária após as diferenciações ( $p\text{-valor} = 0,1$ ), dessa forma, é possível prosseguir com o procedimento de seleção do modelo.

Para tal, se segue a análise dos gráficos ACF e PACF.

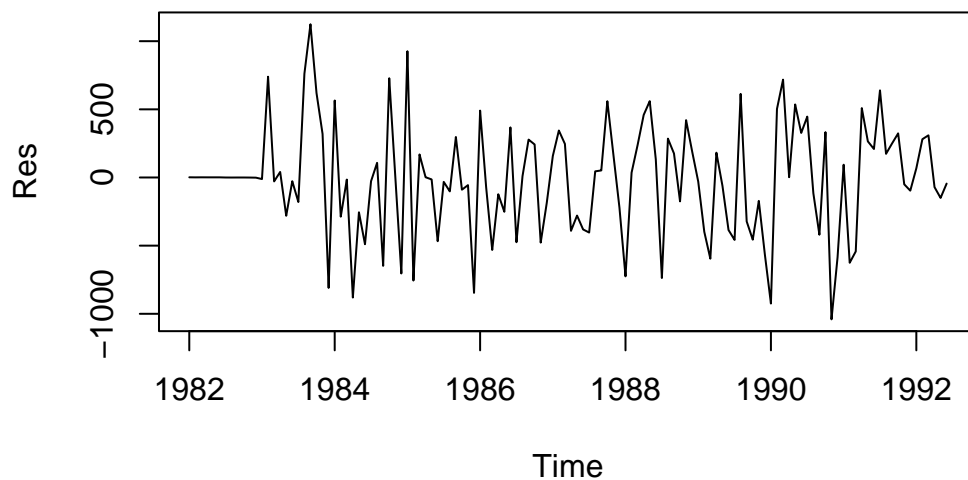


É possível perceber que no ACF há uma quebra no primeiro lag simples, ao passo que ocorre uma quebra no segundo lag do PACF. Porém, nenhum dos gráficos apresenta comportamento bem definido, dificultando a determinação do restante dos parâmetros do modelo SARIMA. Nesse contexto, foi feito a combinação dos valores de  $p$  e  $q$ , entre 0 e 2, e de  $P$  e  $Q$ , nos valores 0 ou 1, desconsiderando quando  $p$  e  $q$  fossem zero simultaneamente e, analogamente, para  $P$  e  $Q$ . O melhor modelo, criado pelas combinações de valores explicitados acima, foi selecionado segundo o menor valor de AICc.

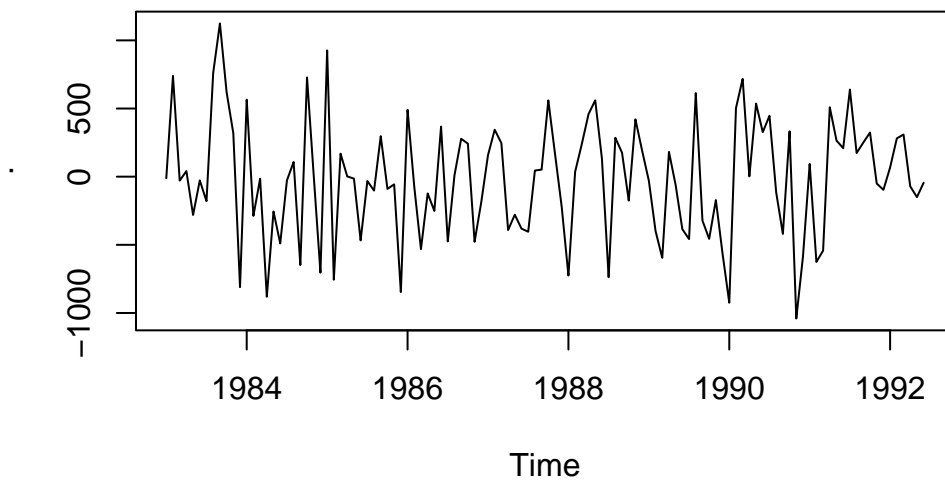
$p = 0$  ,  $d = 1$  ,  $q = 0$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  , AICc = 1750.899  
 $p = 0$  ,  $d = 1$  ,  $q = 1$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  , AICc = 1729.316  
 $p = 0$  ,  $d = 1$  ,  $q = 3$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  , AICc = 1729.225  
 $p = 1$  ,  $d = 1$  ,  $q = 2$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  , AICc = 1729.027  
 $p = 2$  ,  $d = 1$  ,  $q = 0$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  , AICc = 1724.938  
 $p = 0$  ,  $d = 1$  ,  $q = 1$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 1$  , AICc = 1709.211  
 $p = 2$  ,  $d = 1$  ,  $q = 0$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 1$  , AICc = 1706.182

[1] 1706.182

A partir dos AICc, o modelo selecionado foi  $SARIMA(2,1,0)(0,1,1)_{12}$ , assim, pode-se prosseguir com a análise dos resíduos.



É notável que há uma sequência de resíduos iguais a zero no início da série, e estes podem afetar as análises, dessa forma, os zeros foram desconsiderados, sendo feita análise de resíduos de um ano após o início da série, dados pela figura a seguir:



### KPSS Test for Level Stationarity

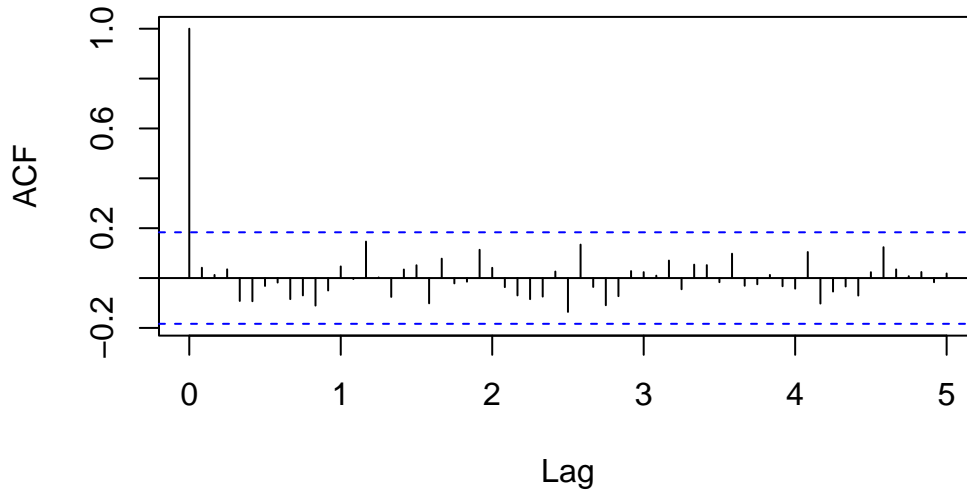
data: E

KPSS Level = 0.062374, Truncation lag parameter = 4, p-value = 0.1

Os resíduos aparentam ser estacionários, com média 0 e com variância constante. Agora, foi testada a estacionariedade com um teste KPSS, cujas as hipóteses já foram mencionadas. Como resultado, obtém-se que o processo é estacionário (p-valor  $> 0, 1$ ).

Além disso, a partir do gráfico do ACF, mostrado abaixo, pode-se verificar a independência dos resíduos.

### Series E

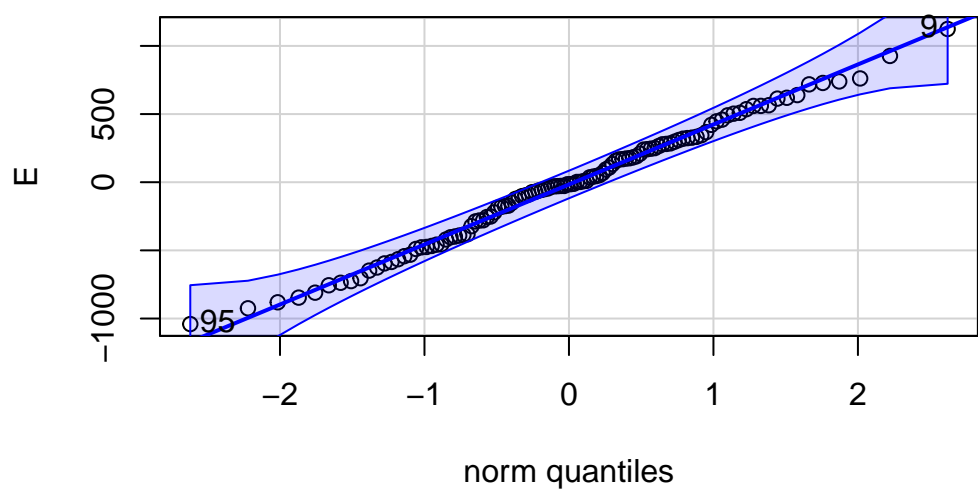
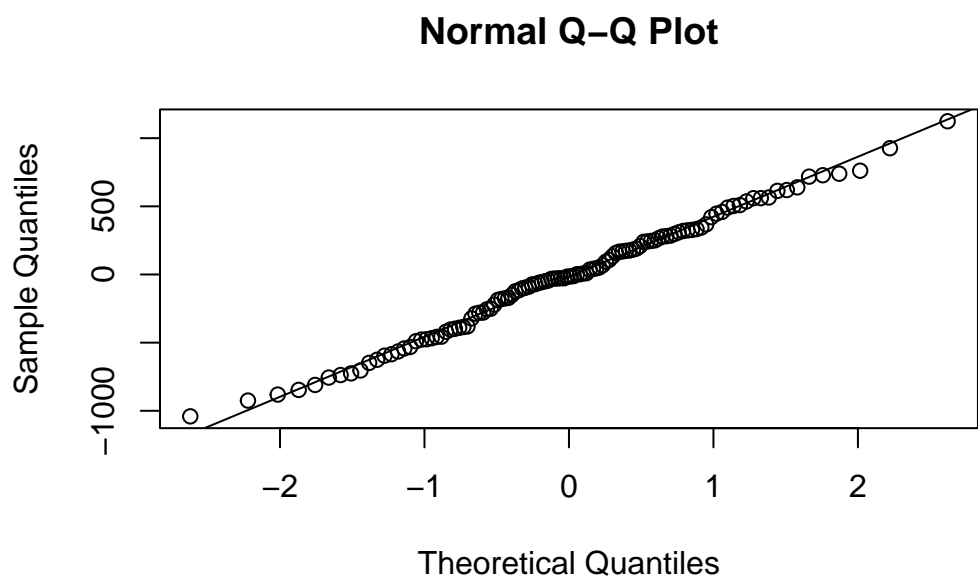


Observamos que praticamente todos os valores das autocorrelações não são significantes, indicando a independência dos resíduos, o que necessita a confirmação pelo teste de Ljung-Box, com base nas hipóteses:

$$\begin{cases} H_0 : \text{Todas as correlações são iguais a zero} \\ H_1 : \text{Ao menos uma correlação é diferente de zero} \end{cases} \quad (2)$$

Assim, utilizando o teste com 20 graus de liberdade, obteve-se p-valor de 0, 8923, e para 15, obteve-se 0, 8741, dessa maneira, utilizando o  $\alpha = 0, 05$ , não se rejeita  $H_0$ , corroborando com a análise gráfica de que os resíduos são independentes. Então, pode-se testar a normalidade, primeiramente, por meio de uma representação gráfica.





[1] 9 95

Shapiro-Wilk normality test

```
data: E
W = 0.99331, p-value = 0.8598
```

Box-Ljung test

```
data: E
X-squared = 12.639, df = 20, p-value = 0.8923
```

```
      ar1      ar2      sma1
-0.5096443 -0.3354277 -0.5080884
```

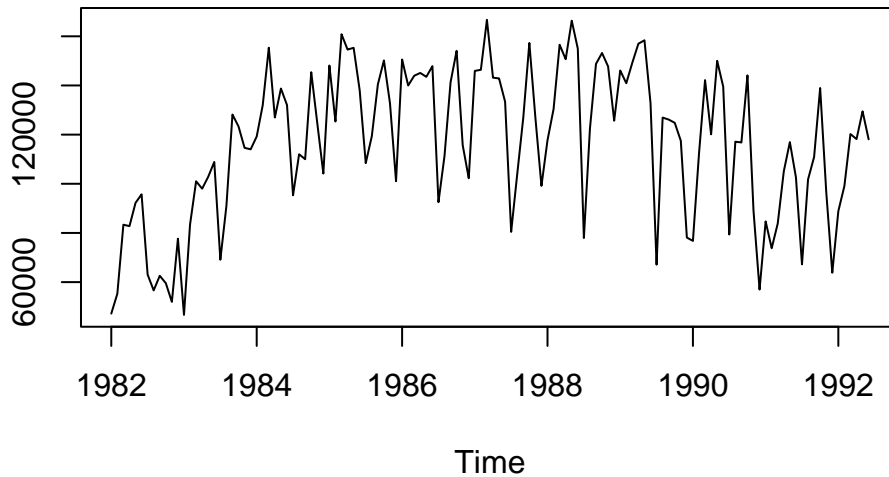
```
[1] 189154.6
```

Assim, pela imagem, com envelope de 95%, é basicamente certa a normalidade dos resíduos, para confirmar, é conduzido o teste de Shapiro-Wilk, sob hipóteses:

$$\begin{cases} H_0 : \text{Os resíduos seguem distribuição normal} \\ H_1 : \text{Os resíduos não seguem distribuição normal} \end{cases} \quad (3)$$

Assim, com nível de significância de 5%, o teste de Shapiro-Wilk confirma que os resíduos estão distribuídos segundo uma distribuição normal.

## Com transformação de Box-Cox



Percebe-se grande similaridade com a série original, e, pode-se prosseguir da mesma maneira que anteriormente com o procedimento de seleção. Primeiramente, testando a estacionariedade da série com o teste KPSS a um nível de significância de 5%, é possível concluir que a série não é estacionária ( $p\text{-valor} = 0,03658$ ). Dessa maneira, analogamente à análise da série sem a transformação Box-Cox, é necessário aplicar 1 diferenciação simples e de uma diferenciação sazonal para tornar a série estacionária, sendo  $d = 1$  e  $D = 1$ . Da mesma maneira que na análise anterior, foi utilizado um sazonal de 12 para a diferenciação sazonal, dessa maneira, a série diferenciada é a que se segue:

### KPSS Test for Level Stationarity

data: x2

KPSS Level = 0.5226, Truncation lag parameter = 4, p-value = 0.03658

[1] 1

[1] 1

Series: x2

ARIMA(0,1,1)(0,1,1)[12]

Coefficients:

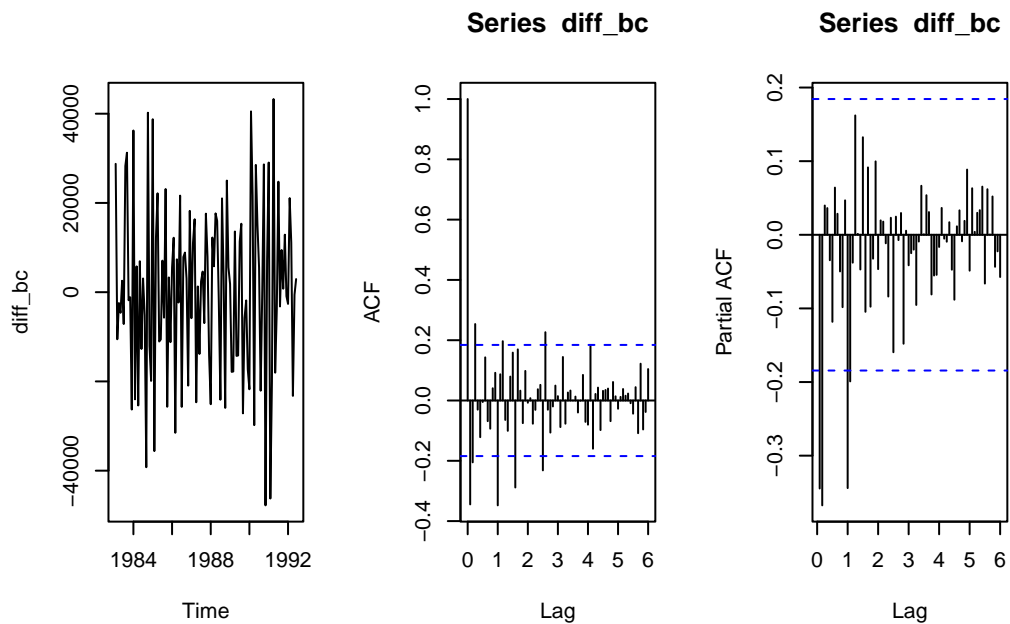
	ma1	sma1
	-0.5166	-0.5532
s.e.	0.0825	0.1060

$\sigma^2 = 238912745$ : log likelihood = -1251.65  
AIC=2509.31 AICc=2509.53 BIC=2517.49

KPSS Test for Level Stationarity

data: diff\_bc

KPSS Level = 0.044088, Truncation lag parameter = 4, p-value = 0.1



Assim ficou a série a partir de 1 diferenciação simples e sazonal. Agora, analogamente à série sem transformação, é possível observar que a série tem aparência estacionária, o que deve ser confirmado novamente a partir do teste de KPSS de sob hipóteses já explicitadas. Assim, a partir de um p-valor maior que 0.1, pôde-se confirmar que a série diferenciada é estacionária e que é possível prosseguir com o procedimento de seleção.

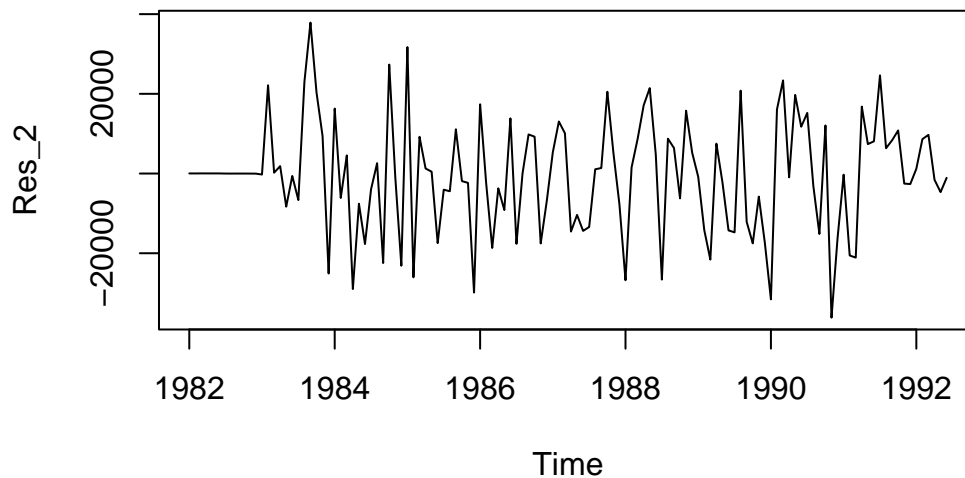
p = 0 , d = 1, q = 0 , P = 0 , D = 1, Q = 0 , AICc = 2553.741  
p = 0 , d = 1, q = 1 , P = 0 , D = 1, Q = 0 , AICc = 2531.436

$p = 0$  ,  $d = 1$  ,  $q = 3$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  ,  $AICc = 2531.137$   
 $p = 1$  ,  $d = 1$  ,  $q = 2$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  ,  $AICc = 2530.789$   
 $p = 2$  ,  $d = 1$  ,  $q = 0$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 0$  ,  $AICc = 2526.922$   
 $p = 0$  ,  $d = 1$  ,  $q = 1$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 1$  ,  $AICc = 2509.529$   
 $p = 2$  ,  $d = 1$  ,  $q = 0$  ,  $P = 0$  ,  $D = 1$  ,  $Q = 1$  ,  $AICc = 2506.5$

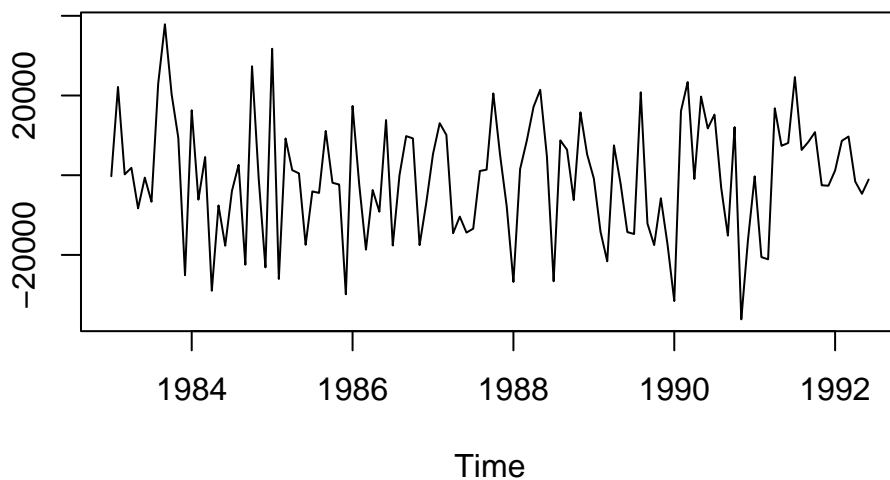
[1] 2506.5

Assim, pode-se observar comportamento similar ao da série sem a transformação de Boxcox. Nesse sentido, no gráfico ACF há uma quebra no primeiro lag simples, ao passo que há uma quebra no segundo lag no gráfico PACF. Nenhum dos gráficos apresentou comportamento bem definido. Para os lags sazonais, percebem-se valores significativos apenas no primeiro lag, sendo assim, os valores de  $p$  e  $q$  serão testados entre 0 e 2, já  $P$  e  $Q$ , nos valores 0 ou 1, desconsiderando quando  $p$  e  $q$  forem 0 simultaneamente, e analogamente para  $P$  e  $Q$ . A partir dos  $AICc$ , observou-se o melhor menor valor de 2506, 5 e assim, o modelo ajustado foi um  $SARIMA(2, 1, 0)(0, 1, 1)_{12}$ , assim com no ajuste de modelo sem utilização da transformação de Box-Cox.

Seguindo para análise dos resíduos, é possível observar que há uma sequência de resíduos iguais a zero no início da série.

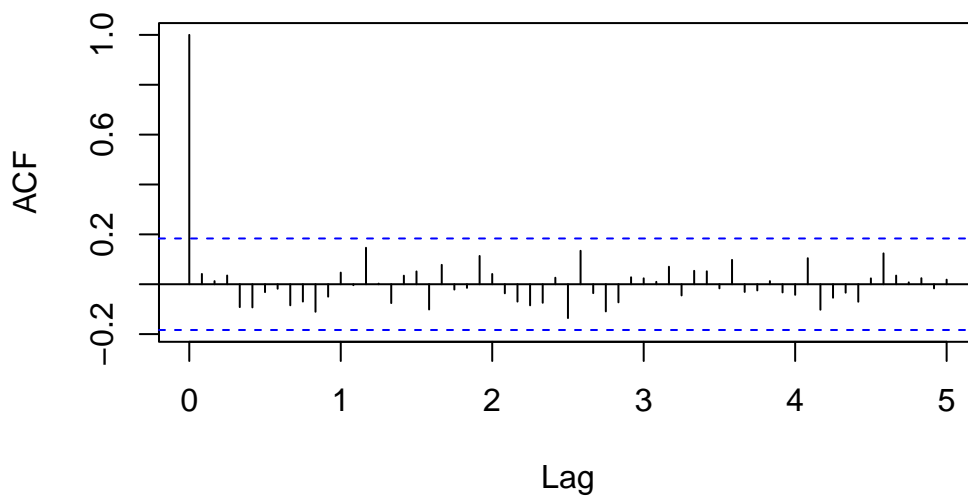


Assim, analogamente à série sem transformação, serão considerados apenas os resíduos um ano após o início das observações, dados pela figura a seguir:

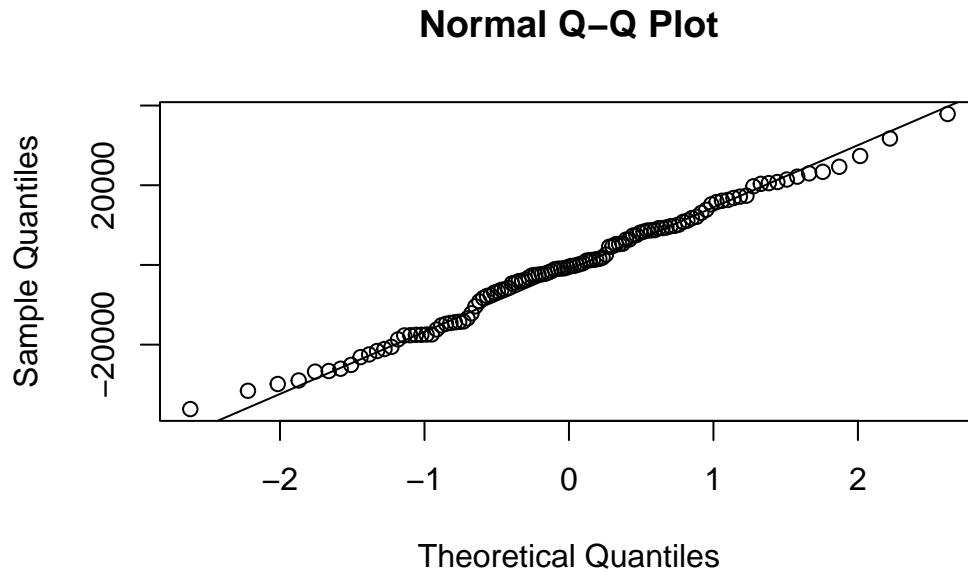


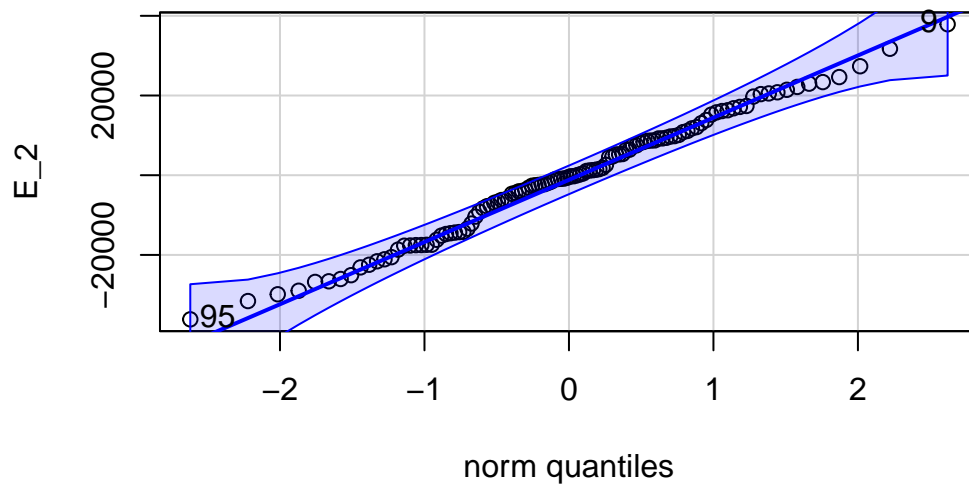
Analisando graficamente, os resíduos aparentam estacionariedade, mas para testar se essa suspeita é correta, utiliza-se o teste de KPSS. O teste mostra que os resíduos são, de fato, estacionários ( $p\text{-valor} > 0, 1$ ). Além disso, pode-se visualizar pelo gráfico do ACF, a independência dos resíduos:

### ACF dos resíduos (Box-Cox)



Assim, de maneira análoga à vista na série sem transformação de Box-Cox, basicamente não há valores significativos, e, para realmente confirmar independência, utiliza-se novamente o teste do tipo Ljung-Box. Assim, com 20 graus de liberdade, o p-valor foi de 0,86, e com 15 graus, o p-valor resultou em 0.8443, portanto, para um nível de significância de 5%, não se rejeita  $H_0$ , indicando que a suspeita após a análise gráfica estava correta, e os resíduos realmente são independentes. Por fim, pode-se prosseguir com a análise da normalidade dos resíduos, por meio da análise gráfica primeiramente.





```
[1] 9 95
```

Shapiro-Wilk normality test

data: E\_2

W = 0.99199, p-value = 0.7513

KPSS Test for Level Stationarity

data: E\_2

KPSS Level = 0.062576, Truncation lag parameter = 4, p-value = 0.1

Box-Ljung test

data: E\_2

X-squared = 13.39, df = 20, p-value = 0.86

ar1	ar2	sma1
-0.5059202	-0.3433980	-0.5264162

```
[1] 224602516
```



A partir da análise gráfica, é possível observar indicação de normalidade dos resíduos. Seguindo com o teste de Shapiro-Wilk, há evidências de que os resíduos seguem distribuição normal (p-valor=0, 7513).

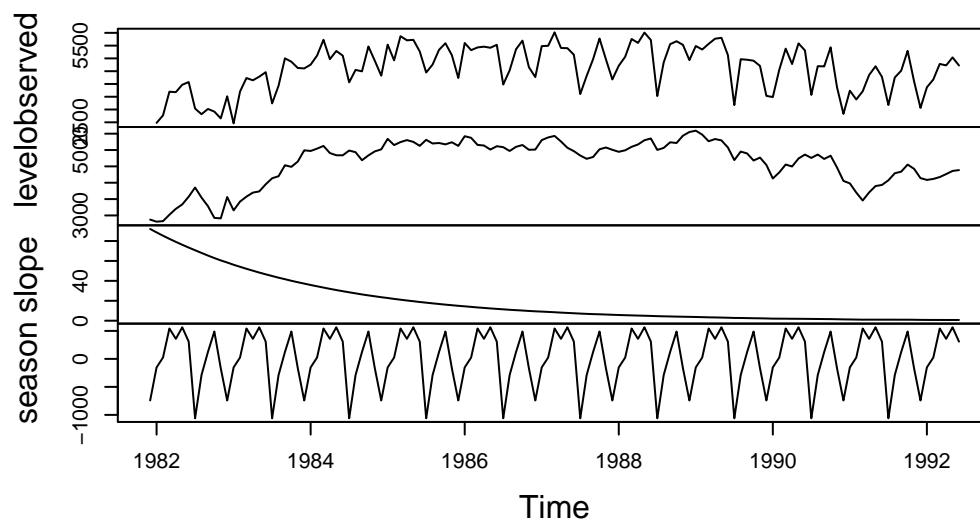
## Modelos ETS: seleção, transformações e resíduos

Modelos ETS tem uma estrutura que permite descrever os modelos de alisamento exponencial em função dos tipos das suas componentes, sendo elas: erro, tendência e sazonalidade. Nesses modelos a tendência pode ser de 5 tipos, sendo: sem tendência, com tendência aditiva, com tendência aditiva mais damped, com tendência multiplicativa e com tendência multiplicativa mais damped. Já a sazonalidade pode ser de 3 tipos: sem sazonalidade, com sazonalidade aditiva e com sazonalidade multiplicativa. Combinar os tipos diferentes de tendência com os tipos diferentes de sazonalidade, resulta em 15 tipos diferentes de modelos que podem ser ajustados. Em adição, o termo de erro pode ser incluído de maneira aditiva ou de maneira multiplicativa, resultando em um total de 30 modelos possíveis.

Modelo	AIC	AICc	BIC
ETS(A,Ad,A)	2154.36	2160.76	2205.42
ETS(A,A,A)	2157.50	2163.17	2205.72
ETS(A,N,A)	2162.35	2166.71	2204.89
ETS(M,Ad,M)	2179.42	2185.81	2230.47
ETS(M,M,M)	2181.29	2187.68	2232.34
ETS(M,A,M)	2184.06	2189.72	2232.27
ETS(M,M,M)	2192.25	2197.92	2240.47
ETS(M,Ad,A)	2194.14	2200.54	2245.20
ETS(M,A,A)	2196.12	2201.79	2244.34
ETS(M,N,M)	2196.41	2200.78	2238.96

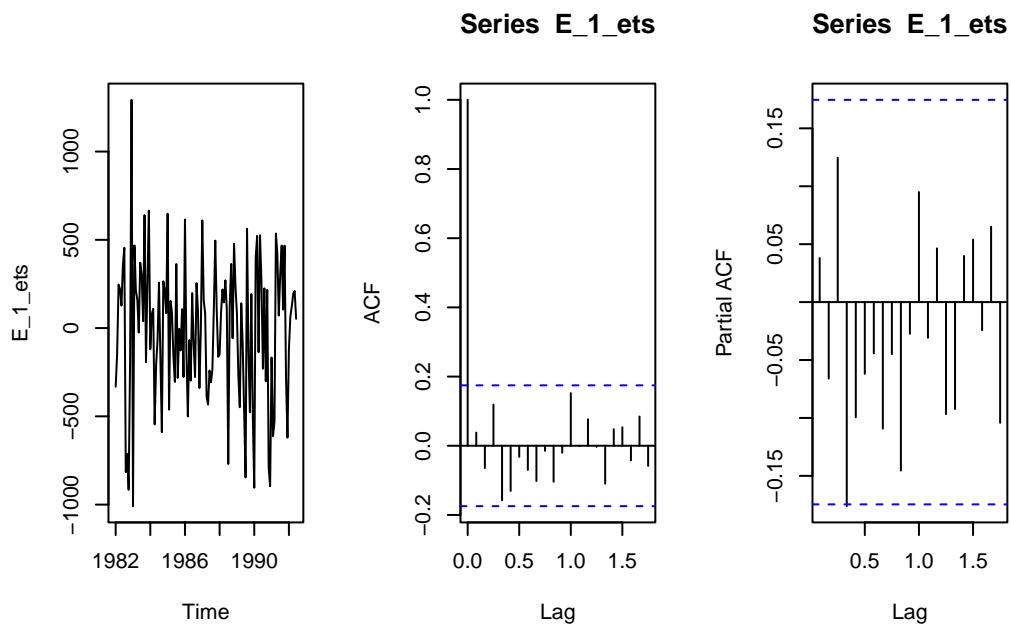
Após varredura dos modelos ETS possíveis, o modelo com menor AICc selecionado foi um ETS(A,  $A_d$ , A), ou seja, um modelo ETS com sazonalidade aditiva, tendência aditiva com damped e erro aditivo. A decomposição da série pode ser observada no gráfico abaixo.

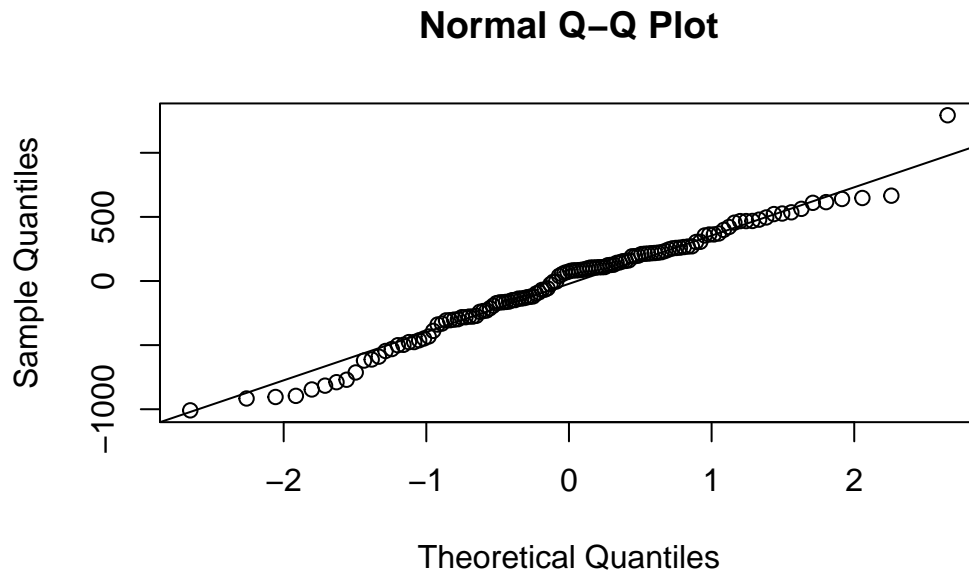
## Decomposition by ETS(A,Ad,A) method



## Resíduos

Seguindo com a análise de resíduos do modelo ajustado, observa-se os seguintes resultados.





Analisando graficamente, os resíduos aparentam estacionariedade. Aplicando o teste de KPSS é possível observar que, de fato, os resíduos são estacionários ( $p\text{-valor} > 0, 1$ ). Além disso, pode-se visualizar pelo gráfico do ACF, a independência dos resíduos.

#### KPSS Test for Level Stationarity

```
data: E_1_ets
KPSS Level = 0.064386, Truncation lag parameter = 4, p-value = 0.1
```

#### Box-Ljung test

```
data: E_1_ets
X-squared = 16.069, df = 15, p-value = 0.3775
```

#### Shapiro-Wilk normality test

```
data: E_1_ets
W = 0.98013, p-value = 0.06068
```

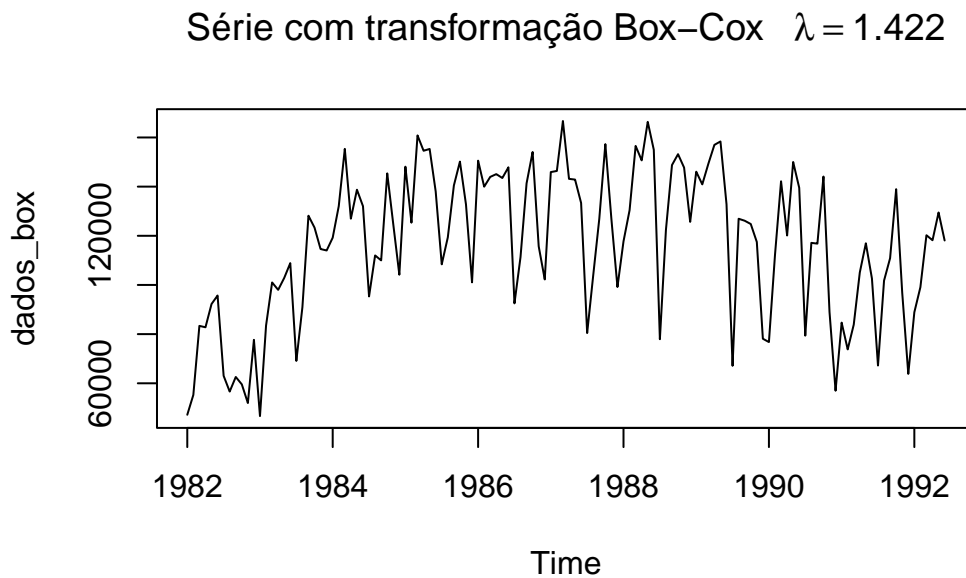
	Estatística	p_valor
W	0.9801266	0.0606769
KPSS Level	0.0643859	0.1000000
X-squared	16.0693204	0.1880817

Seguindo para os teste de independência e normalidade dos resíduos, é possível concluir que eles são independentes (p-valor = 0,188) e estão normalmente distribuídos (p-valor = 0,061), a um nível de significância de 5%. No entanto, especificamente para o teste de normalidade, não se pode dizer que é um teste muito confiável, já que alterar o nível de significância pode alterar o resultado do teste.

## Modelo com transformação

### Seleção

Para a seleção de um modelo utilizando a transformação de Box-Cox, primeiramente sou definido o valor de  $\lambda$  que melhor se ajustasse a série, sendo esse valor de  $\lambda = 1,422$ , como pode ser observado no gráfico abaixo.



Seguindo o mesmo procedimento aplicado para selecionar o modelo sem transformação de Box-Cox, foi feita uma varreura de todos os modelos possíveis e o melhor deles foi selecionado segundo o critério AICc.

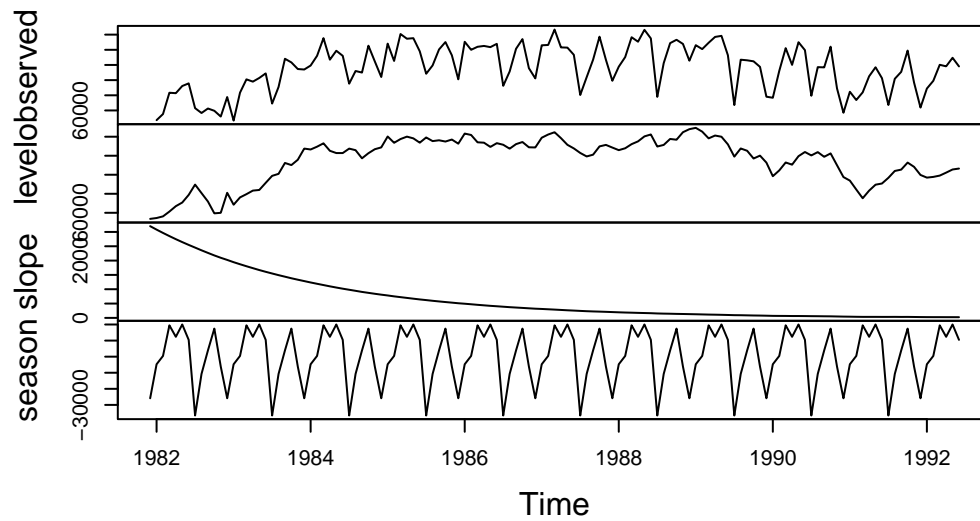
Modelo transformado	AIC	AICc	BIC
ETS(A,Ad,A)	3044.31	3050.71	3095.37
ETS(A,A,A)	3047.45	3053.12	3095.67
ETS(A,N,A)	3049.98	3054.35	3092.53
ETS(M,Ad,M)	3072.16	3078.55	3123.21

(continued)

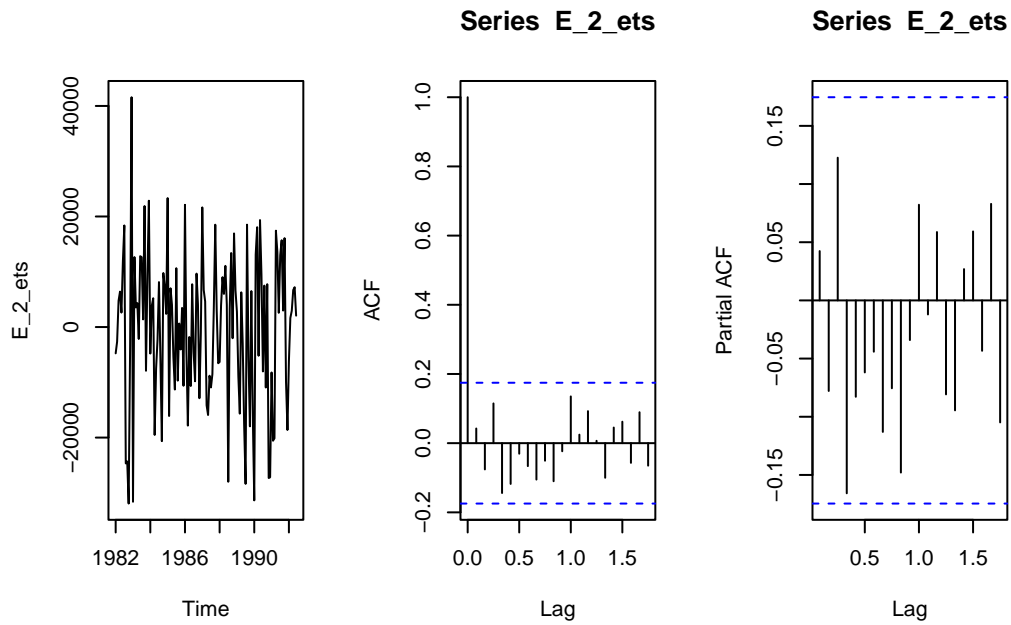
Modelo transformado	AIC	AICc	BIC
ETS(M,M,M)	3073.43	3079.82	3124.49
ETS(M,A,M)	3079.96	3085.63	3128.18
ETS(M,M,M)	3086.13	3091.80	3134.35
ETS(M,N,M)	3087.88	3092.24	3130.42
ETS(M,Ad,A)	3101.83	3108.22	3152.88
ETS(M,A,A)	3108.98	3114.65	3157.20

O modelo selecionado segundo o menor AICc, levando em consideração a transformação de Box-Cox, foi o mesmo selecionado anteriormente, sendo  $ETS(A, A_d, A)$ , ou seja, um modelo ETS com sazonalidade aditiva, tendência aditiva com damped e erro aditivo.

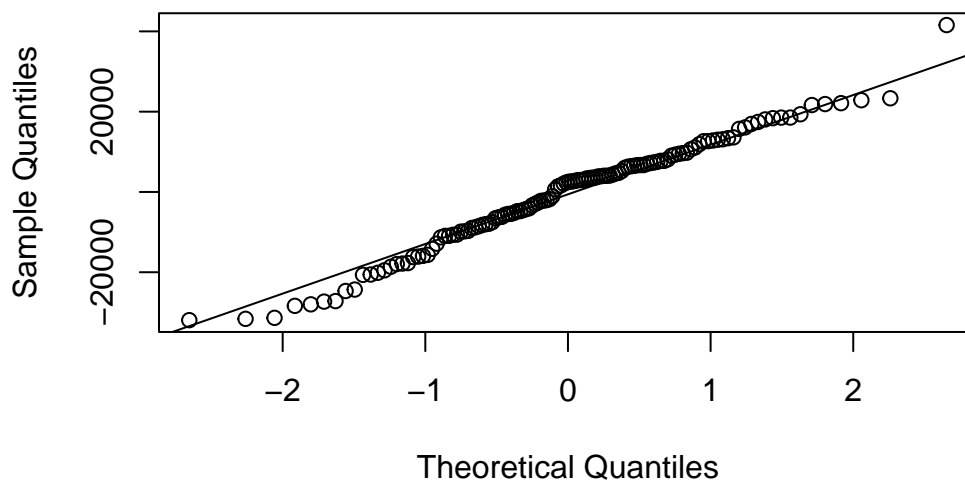
### Decomposition by ETS(A,Ad,A) method



## Resíduos



## Normal Q-Q Plot



Analisando graficamente, os resíduos aparentam estacionariedade. Aplicando o teste de KPSS é possível observar que, de fato, os resíduos são estacionários ( $p\text{-valor} > 0, 1$ ). Além disso,

pode-se visualizar pelo gráfico do ACF, a independência dos resíduos.

#### KPSS Test for Level Stationarity

data: E\_2\_ets

KPSS Level = 0.071953, Truncation lag parameter = 4, p-value = 0.1

#### Box-Ljung test

data: E\_2\_ets

X-squared = 15.575, df = 15, p-value = 0.4109

#### Shapiro-Wilk normality test

data: E\_2\_ets

W = 0.98338, p-value = 0.1257

	Estatística	p_valor
W	0.9833828	0.1256580
KPSS Level	0.0719535	0.1000000
X-squared	15.5745517	0.2115083

Seguindo para os teste de independência e normalidade dos resíduos, é possível concluir que eles são independentes (p-valor = 0,4109) e estão normalmente distribuídos (p-valor = 0,1257), a um nível de significância de 5%. Nesse caso, tanto o teste de independência quanto o teste de normalidade tem resultados mais confiáveis.

## Estudo de desempenho preditivo

### Resultados da Janela Deslizante

#### Performance em relação aos horizontes de previsão

##### ARIMA

##### ETS

## Resultados

apresente em tabelas e gráficos as previsões dos 4 modelos selecionados e também apresente em uma tabela os resultados de acurácia dos 4 modelos selecionados e dos modelos benchmarks.

Comente os resultados de modo objetivo;

## **Apêndice**

Todo o projeto de composição deste documento pode ser encontrado aqui: [https://github.com/cesar-galvao/trabalhos\\_series](https://github.com/cesar-galvao/trabalhos_series)