

Análise de Séries Temporais - Trabalho 2

Davi Guerra Alves - Henrique Oliveira Dumay

2023-07-02

Apresentação

A série analisada consiste na série número 1686 pertencente ao banco de dados da competição de previsão M3, disponível no pacote *Mcomp* do software R. A série descreve o número de carregamentos com código *TD-AUTOUNITS*, mensalmente, de outubro de 1984 a setembro de 1993.

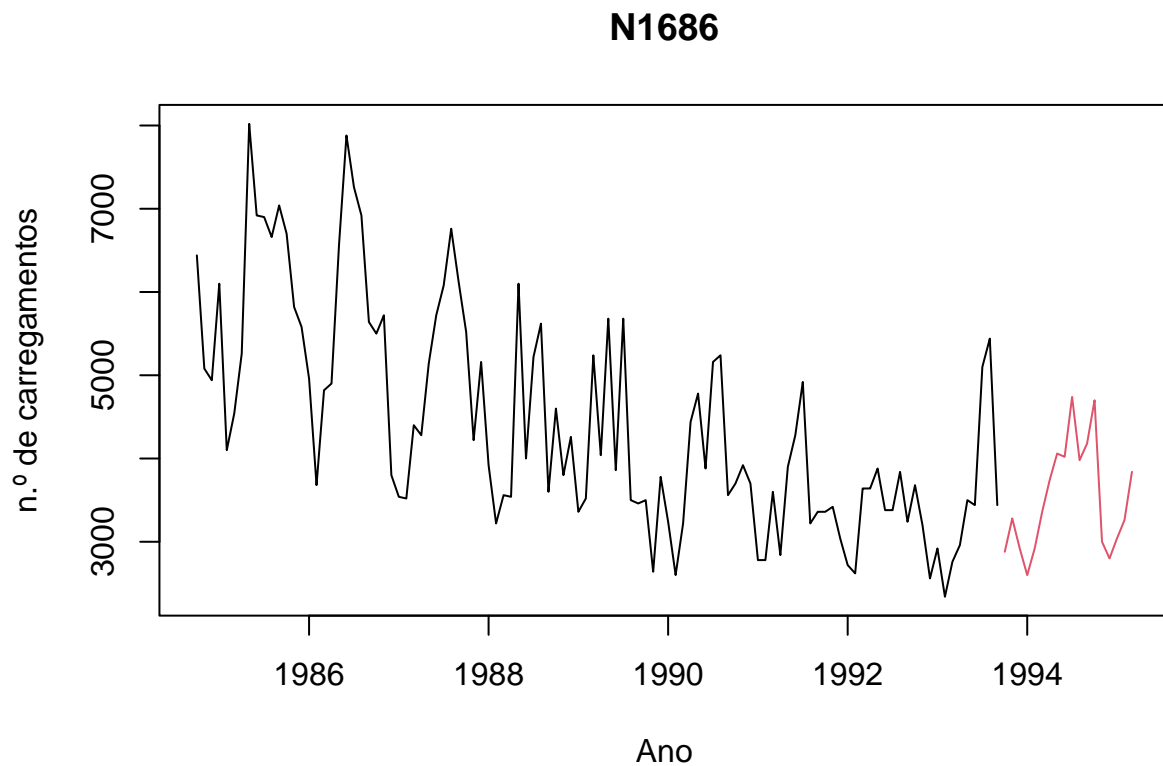


Figure 1: Comportamento da série ao longo do tempo

Decomposição MSTL

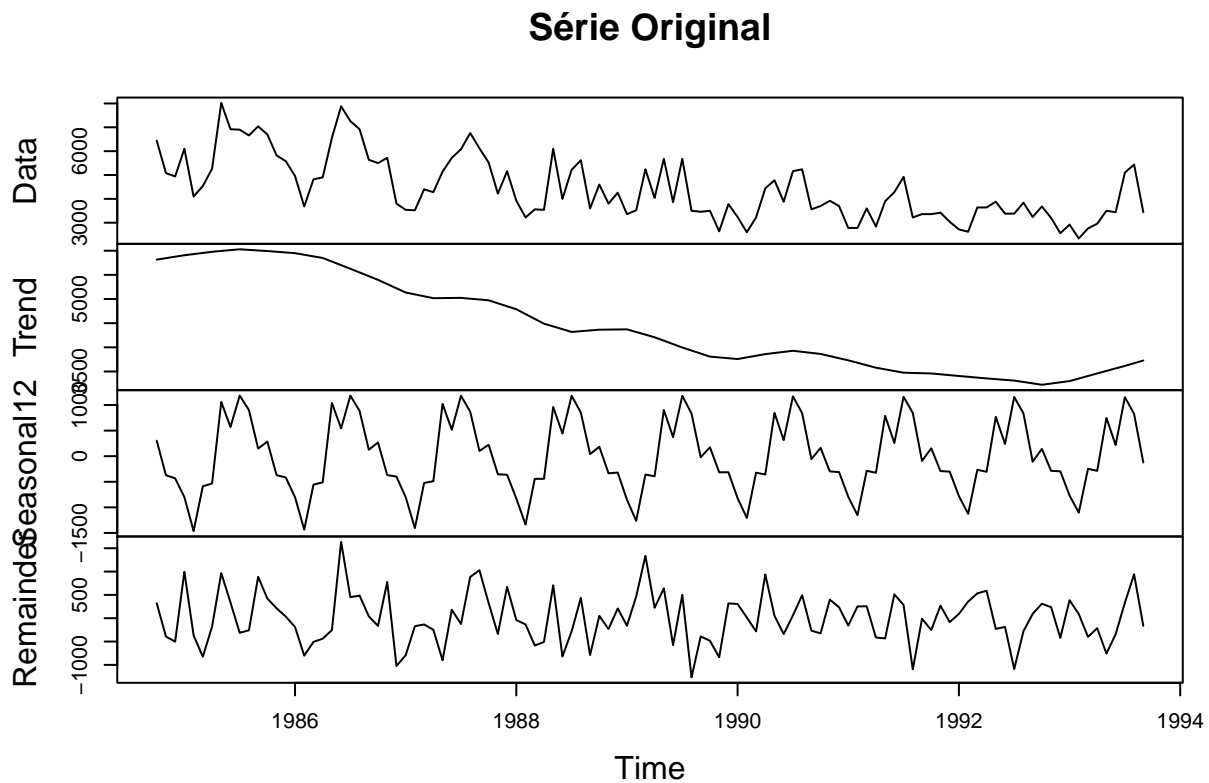


Figure 2: Decomposição MSTL

A decomposição MSTL mostra os componentes de tendência, sazonalidade e erro da série estudada. Percebe-se a presença de uma tendência crescente, com múltiplas sazonalidades que apresentam mudança do comportamento ao longo do tempo. É possível observar, graficamente, um alargamento da sazonalidade ao fim da série quando comparando ao início da série.

Modelos ARIMA

A presença do componente de tendência explicita a não-estacionariedade da série original. A função `ndiffs()` é utilizada para estimar o número de diferenças exigidas para tornar a série estacionária por meio de um teste de raiz unitária, com a hipótese nula de que a série tem raízes estacionárias contra a hipótese alternativa de que a série tem raiz unitária. O teste retorna o menor número de diferenças exigidas para o teste em um nível de significância de 95%. Já a função `nsdiffs()` utiliza testes de raiz unitária para determinar o número de diferenças sazonais para tornar a série estacionária.

Com o uso das funções acima, obteve-se o valor para $d = 1$ e $D = 0$. Os modelos candidatos terão a forma:

$$SARIMA(p, 1, q) \times (P, 0, Q)_{12}$$

A estacionariedade da série pode ser testada utilizando o teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS), com a hipótese nula de que a série é estacionária. O teste resulta em um valor de 0.0221285, com p-valor de 0.1, que não nos permite rejeitar a hipótese nula a um nível de significância $\alpha = 0,05$.

Consideramos que a série é, agora, estacionária, observamos os gráficos da função de autocorrelação (ACF) e da função de autocorrelação parcial (PACF) em busca de possíveis autocorrelações entre os diferentes atrasos da série. Os gráficos a seguir ilustram a série diferenciada, assim como os gráficos das funções de ACF e PACF.

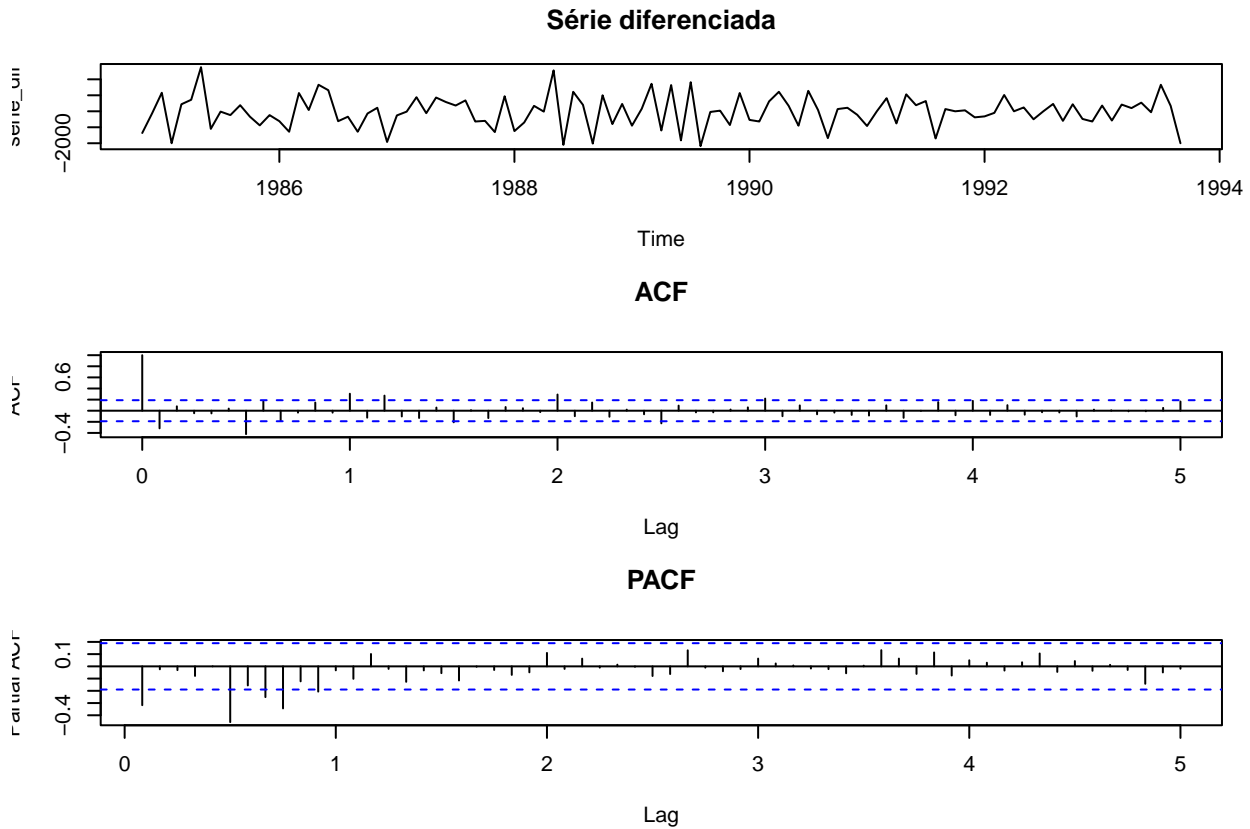


Figure 3: Gráficos ACF e PACF

Dos gráficos apresentados, pode-se afirmar que a série diferenciada não apresenta um padrão claro de autocorrelações simples e sazonais que permita inferir diretamente a modelagem. Neste sentido, serão testados valores diferentes para p , P , q e Q e os diferentes modelos serão comparados por meio do critério AIC.

Para os diferentes valores de (p, q, P, Q) teremos:

```
## p = 0 , q = 0 , P = 0 , Q = 0 , AICc = 1782.086
## p = 0 , q = 1 , P = 0 , Q = 0 , AICc = 1772.446
## p = 0 , q = 3 , P = 0 , Q = 0 , AICc = 1765.736
## p = 2 , q = 3 , P = 0 , Q = 0 , AICc = 1745.548
## p = 0 , q = 2 , P = 1 , Q = 0 , AICc = 1745.396
## p = 1 , q = 1 , P = 1 , Q = 0 , AICc = 1744.079
## p = 2 , q = 2 , P = 1 , Q = 0 , AICc = 1743.72
## p = 3 , q = 3 , P = 1 , Q = 0 , AICc = 1741.292
## p = 0 , q = 1 , P = 1 , Q = 1 , AICc = 1739.715
```

```
## p = 0 , q = 2 , P = 1 , Q = 1 , AICc = 1735.097
## p = 1 , q = 1 , P = 1 , Q = 1 , AICc = 1733.713
```

O modelo com menor AICc foi o $SARIMA(1, 1, 1) \times (1, 0, 1)_{12}$.

Os coeficientes do modelo proposto, portanto, serão obtidos do cálculo da função Arima com o modelo acima proposto. Os coeficientes do modelo terão a seguinte forma:

$$\phi_1 = 0,3152; \theta_1 = -0,9218; \varphi = 0,9606; \vartheta = -0,7359$$

Para o modelo ARIMA utilizando transformação Box-Cox, os valores para d e D utilizando as funções $ndiffs()$ e $nsdiffs()$ são, respectivamente, 1 e 0. De acordo com o teste KPSS, não se pode rejeitar a hipótese nula da série transformada ser estacionária (KPSS = 0.021624 e p-valor = 0.1). O valor do parâmetro λ da transformação de Box-Cox é 0.09559902.

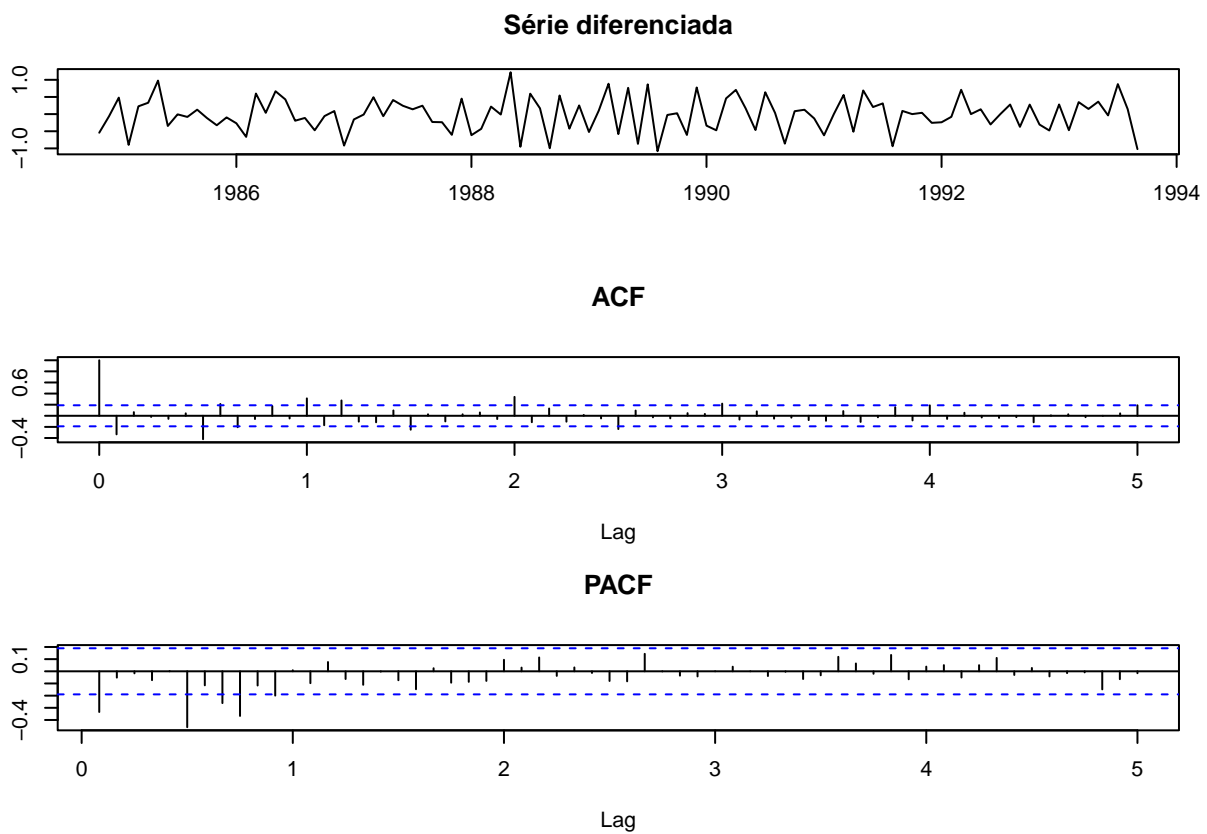


Figure 4: Gráficos ACF e PACF da série transformada (BoxCox)

A análise gráfica das funções de autocorrelação e autocorrelação da série transformada não permite a inferência a respeito de seus parâmetros de maneira clara. Procedeu-se a pesquisa de valores de p , q , P e Q de forma manual de acordo com o critério de AICc.

```
## p = 0 , q = 0 , P = 0 , Q = 0 , AICc = 153.6586
## p = 0 , q = 1 , P = 0 , Q = 0 , AICc = 142.2782
## p = 0 , q = 3 , P = 0 , Q = 0 , AICc = 136.6817
## p = 1 , q = 3 , P = 0 , Q = 0 , AICc = 136.3155
```

```

## p = 2 , q = 3 , P = 0 , Q = 0 , AICc = 107.1571
## p = 0 , q = 1 , P = 1 , Q = 1 , AICc = 103.1689
## p = 0 , q = 2 , P = 1 , Q = 1 , AICc = 99.51568
## p = 1 , q = 1 , P = 1 , Q = 1 , AICc = 98.80396
## p = 3 , q = 3 , P = 1 , Q = 1 , AICc = 95.4485

```

Para a série transformada, temos que o menor nível de AICc foi encontrado com os parâmetros $p = 3$, $q = 3$, $P = 1$ e $Q = 1$, cujo modelo pode ser descrito por $SARIMA(3, 1, 3) \times (1, 0, 1)_{12}$, com coeficientes iguais a:

$\phi_1 = 1,2866$; $\phi_2 = -1,0479$; $\phi_3 = 0,2771$; $\theta_1 = -2,0555$; $\theta_2 = 2,0480$; $\theta_3 = -0,9371$; $\varphi = 0,9995$; $\vartheta = -0,9649$

Análise de Resíduos

Os resíduos do modelo ARIMA sem transformação apresentam o seguinte comportamento gráfico:

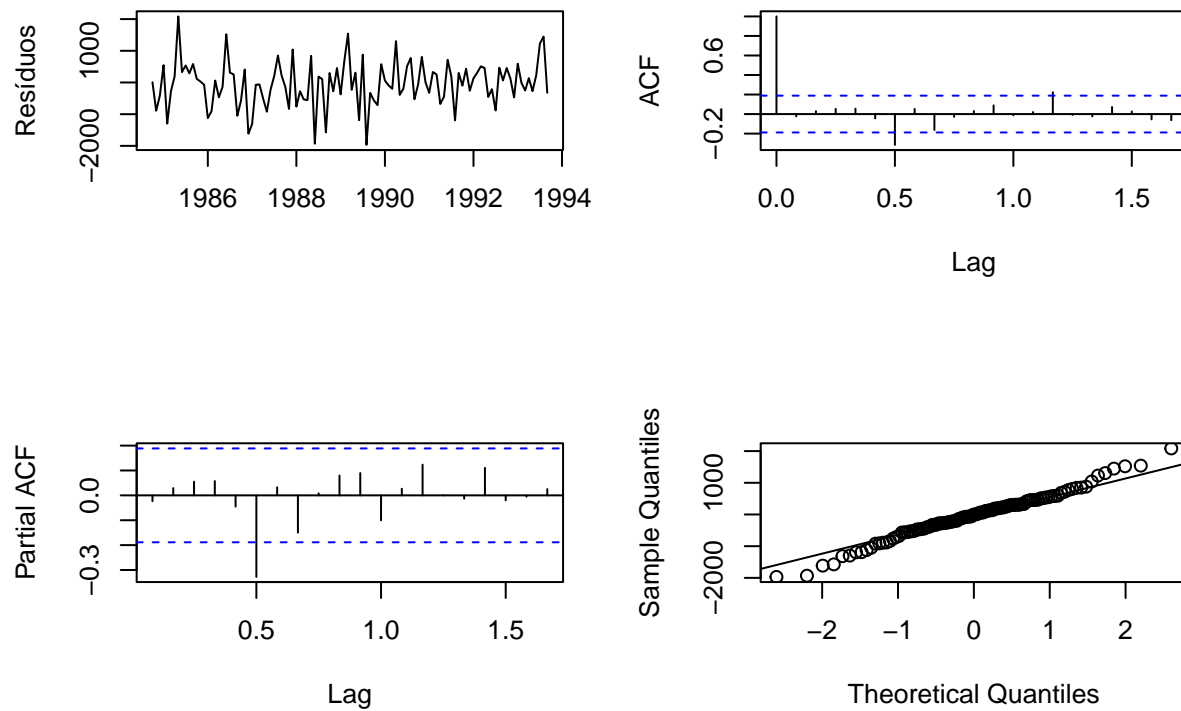


Figure 5: Resíduos ARIMA sem transformação

Já os resíduos do modelo ARIMA com transformação Box-Cox apresentam o seguinte comportamento gráfico:

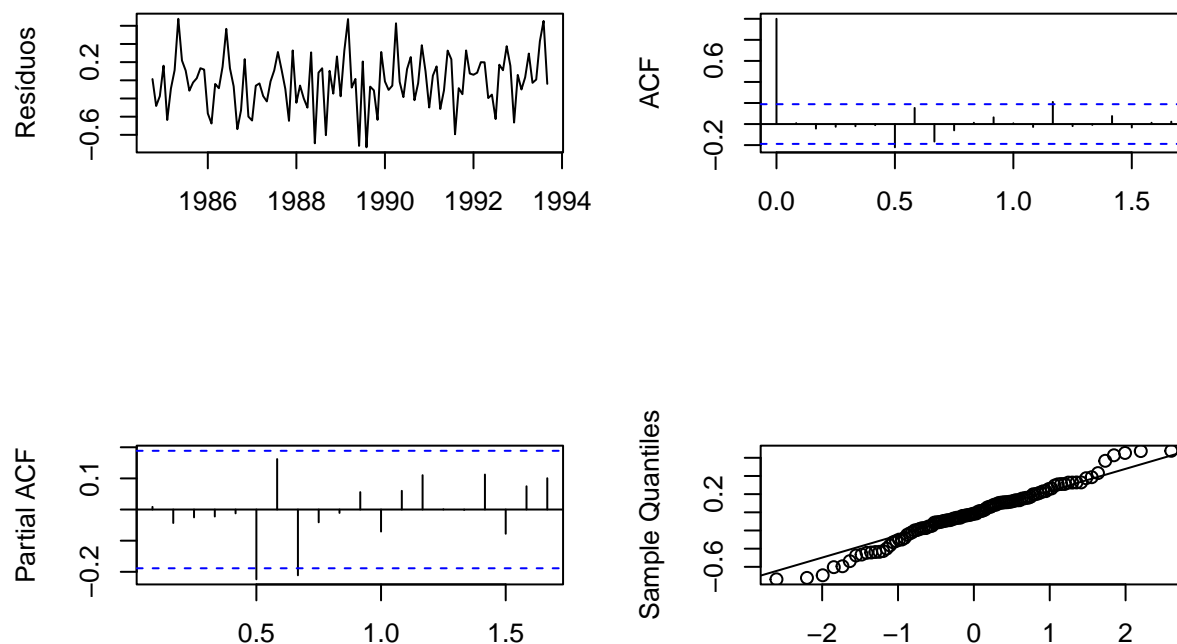


Figure 6: Resíduos ARIMA com transformação boxcox

Graficamente, observa-se que os resíduos de ambos os modelos parecem distribuir-se simetricamente ao redor da origem e não apresentam autocorrelações bem definidas. Precisa-se, entretanto, testá-los para estacionariedade, independência e distribuição normal. Essas hipóteses serão testadas conforme se segue, todas assumindo nível de significância $\alpha = 0.05$.

A estacionariedade será testada a partir do teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS), com a hipótese nula de que a série é estacionária. O teste para o modelo $SARIMA(1, 1, 1) \times (1, 0, 1)_{12}$ e para o mesmo modelo, utilizando a transformação de Box-Cox:

Modelo	KPSS	P-valor
SARIMA sem Box-Cox	0.2133231	0.1
SARIMA com Box-Cox	0.2498156	0.1

De acordo com o teste KPSS, não se pode rejeitar a hipótese de estacionariedade dos resíduos de ambos os modelos.

O teste de independência dos resíduos é realizado a partir do teste Ljung-Box, com a hipótese H_0 de que os resíduos são idenpendentemente distribuídos. O teste apresenta os seguintes valores para os dois modelos:

Modelo	Chi-Quadrado	Graus de liberdade	P-valor
SARIMA sem Box-Cox	23.72163	15	0.069974

Modelo	Chi-Quadrado	Graus de liberdade	P-valor
SARIMA com Box-Cox	18.7538527	15	0.2251056

Os resultados acima mostram que a independência dos resíduos não pode ser rejeitada ao nível de significância de 5% em ambos os modelos.

A normalidade dos resíduos é testada com o teste Shapiro-Wilk de Normalidade, com H_0 de que os resíduos apresentam distribuição normal. O valor do teste estatístico para os dois modelos trabalhados é:

Modelo	W	P-valor
SARIMA sem Box-Cox	0.986694	0.3620588
SARIMA com Box-Cox	0.9846901	0.2528724

Do resultado acima, não se pode rejeitar a hipótese de normalidade dos resíduos de ambos os modelos, com nível de significância de 5%.

Previsões

As previsões dos modelos:

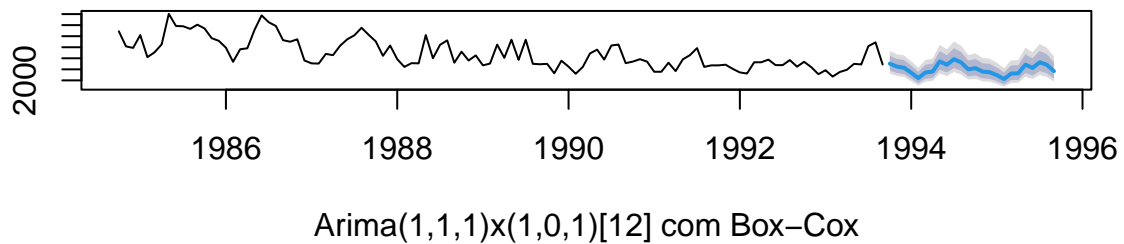
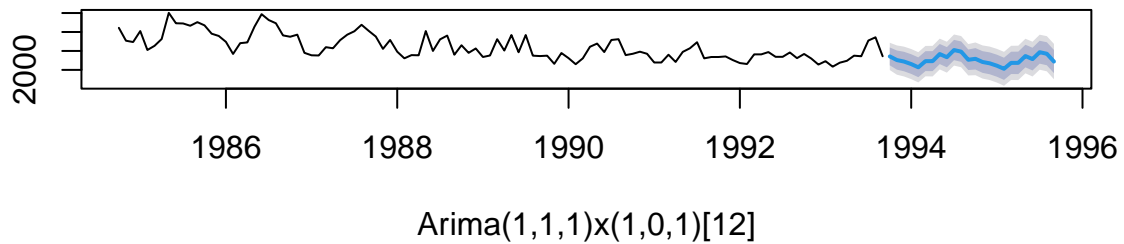


Figure 7: Previsões dos modelos ARIMA

Modelos ETS

O modelo ETS (Error, trend and seasonal) permite descrever os modelos de alisamento exponencial em função dos tipos de suas componentes: tendência, sazonalidade e erro. O modelo utiliza três caracteres como identificação de acordo com a terminologia adotada por Hyndman et al. (2002) e Hyndman et al. (2008). A primeira letra se refere ao componente do erro; a segunda, ao componente da tendência e a terceira, da sazonalidade. A série anteriormente descrita apresenta tendência e sazonalidade claras à decomposição realizada e, portanto, trabalharemos com componentes de modelagem que contenham essas características.

Os modelos que apresentam as características observadas na decomposição e seus respectivos AICcs, considerando um modelo não transformado e um modelo com transformação de Box-Cox, estão representados na tabela abaixo:

Parâmetros	AICc Modelo sem Box-Cox	AICc Modelo com Box-Cox
AAA	1950.568	290.3203
AAA Dumped	1953.429	295.7952
MAA	1947.757	288.9366
MAA Dumped	1942.956	297.1602
MAM	1925.053	292.4207
MAM Dumped	1928.537	291.7428
MMM	1923.268	300.8768
MMM Dumped	1928.128	293.2297

O modelo com menor AICc e, portanto, o modelo com melhor desempenho comparativo, para o modelo não transformado é o “MMM”, em que apresenta componentes multiplicativos para previsões de erros, tendência e sazonalidade. Já para o modelo com transformação de Box-Cox, o modelo com menor AICc é o “MAM” com *dumped*.

O modelo selecionado sem transformação apresenta a seguinte estrutura:

$$\begin{aligned}
 \mu_t &= l_{t-1}b_{t-1}s_{t-m} \\
 l_t &= l_{t-1}b_{t-1} + \frac{\alpha\epsilon_t}{s_{t-m}} \\
 b_t &= b_{t-1} + \frac{\beta\epsilon_t}{s_{t-m}l_{t-1}} \\
 s_t &= s_{t-m} + \frac{\gamma\epsilon_t}{l_{t-1}b_{t-1}}
 \end{aligned}$$

Enquanto o modelo selecionado para o conjunto de dados transformados por Box-Cox tem a seguinte estrutura:

$$\begin{aligned}
 \mu_t &= (l_{t-1} + \phi b_{t-1})s_{t-m} \\
 l_t &= (l_{t-1} + \phi b_{t-1})(1 + \alpha\epsilon_t) \\
 b_t &= \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1})\epsilon_t \\
 s_t &= s_{t-m}(1 + \gamma\epsilon_t)
 \end{aligned}$$

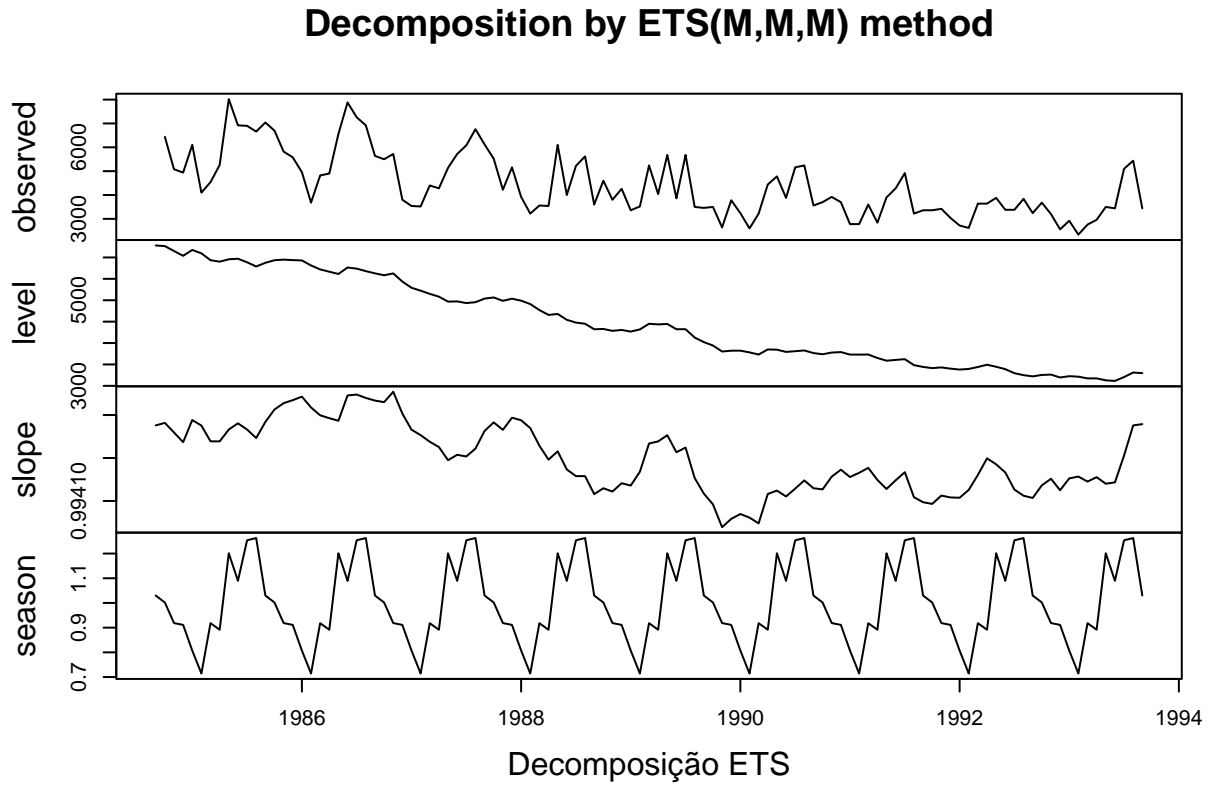


Figure 8: Decomposição ETS

Utilizando o modelo ETS com uma transformação do tipo Box-Cox

O modelo selecionado apresenta a seguinte estrutura:

$$\begin{aligned}\mu_t &= l_{t-1}b_{t-1}s_{t-m} \\ l_t &= l_{t-1}b_{t-1} + \frac{\alpha\epsilon_t}{s_{t-m}} \\ b_t &= b_{t-1} + \frac{\beta\epsilon_t}{s_{t-m}l_{t-1}} \\ s_t &= s_{t-m} + \frac{\gamma\epsilon_t}{l_{t-1}b_{t-1}}\end{aligned}$$

Resíduos

Os resíduos do modelo ETS sem transformação selecionado são ilustrados na figura a seguir:

`\begin{figure}[H]`

`\{\centering \includegraphics{trabalho2_files/figure-latex/unnamed-chunk-20-1}`

`\}`

\caption{Análise de Resíduos do ETS}\label{fig:unnamed-chunk-20}

\end{figure}

Enquanto os gráficos para o modelo ETS selecionado usando transformação de Box-Cox são ilustrados abaixo

\begin{figure}[H]

{\centering \includegraphics{trabalho2_files/figure-latex/unnamed-chunk-21-1}}

}

\caption{Análise de Resíduos do ETS do modelo com Box-Cox}\label{fig:unnamed-chunk-21}

\end{figure}

Observa-se que os resíduos parecem comportar-se de maneira aleatória, com distribuição normal e sem aut

Teste	MMM sem dumped	p-valor	MAM com dumped	Box-Cox	p-valor
:---:	:---:	:---:	:---:	:---:	:---:
KPSS	0.12301	0.1	0.2145	0.1	
Ljung-Box	25.522	0.038	41.505	0.003	
Shapiro-Wilk	0.98751	0.416	0.99443	0.944	

Agora analisando o modelo ETS(M,M,M), com a transformação boxcox, temos os seguintes resultados:

\begin{figure}[H]

{\centering \includegraphics{trabalho2_files/figure-latex/unnamed-chunk-23-1}}

}

\caption{Análise de Resíduos do ETS com boxcox}\label{fig:unnamed-chunk-23}

\end{figure}

Os resultados acima apresentados sugerem que os resíduos do modelo são estacionários, apresentam distri

Uma previsão do modelo para 12 novos períodos é ilustrada abaixo:

\begin{figure}[H]

{\centering \includegraphics{trabalho2_files/figure-latex/unnamed-chunk-24-1}}

}

\caption{Previsão ETS}\label{fig:unnamed-chunk-24}

\end{figure}

\begin{figure}[H]

```
{\centering \includegraphics{trabalho2_files/figure-latex/unnamed-chunk-25-1}}
```

```
}
```

```
\caption{Previsão ETS com transformação boxcox}\label{fig:unnamed-chunk-25}
\end{figure}
```

```
# Estudo de desempenho preditivo por janela deslizando
```

Considerante os dois modelos ARIMA e os dois modelos ETS anteriormente escolhidos, será realizado um estudo de desempenho preditivo por janela deslizando.

```
```r
```

```
Definindo as funções de previsão:
```

```
f_arima <- function(y, h){
```

```
 fit = Arima(y, order=c(1,1,1), seasonal=c(1,0,1))
```

```
 forecast(fit, h)
```

```
}
```

```
f_arima_boxcox <- function(y, h){
```

```
 fit = Arima(y, order=c(3,1,3), seasonal=c(1,0,1), lambda = "auto")
```

```
 forecast(fit, h)
```

```
}
```

```
f_ets <- function(y, h){
```

```
 fit = ets(y, model="MMM", dumped = FALSE)
```

```
 forecast(fit, h)
```

```
}
```

```
f_ets_boxcox <- function(y, h){
```

```
 fit = ets(y, model="MAM", dumped = TRUE, lambda = "auto")
```

```
 forecast(fit, h)
```

```
}
```

```
Calculo dos erros de previsão com a janela flutuante
```

```
CV_arima = tsCV(y=serie,
```

```
 forecastfunction=f_arima,
```

```
 h=5,
```

```
 init = length(serie)-14)
```

```
CV_arima_boxcox = tsCV(y=serie,
```

```
 forecastfunction=f_arima_boxcox,
```

```
 h=5,
```

```
 init = length(serie)-14)
```

```
CV_ets = tsCV(y=serie,
```

```
 forecastfunction=f_ets,
```

```
 h=5,
```

```
 init = length(serie)-14)
```

```
CV_ets_boxcox = tsCV(y=serie,
```

```
 forecastfunction=f_ets_boxcox,
```

```
 h=5,
```

```
 init = length(serie)-14)
```

```
CV_arima %>% round(2) %>% tail(20)
```

```
h=1 h=2 h=3 h=4 h=5
Feb 1992 NA NA NA NA NA
```

```
Mar 1992 NA NA NA NA NA
Apr 1992 NA NA NA NA NA
May 1992 NA NA NA NA NA
Jun 1992 NA NA NA NA NA
Jul 1992 NA NA NA NA NA
Aug 1992 69.94 493.20 239.34 -392.34 431.26
Sep 1992 467.87 228.99 -399.43 426.13 47.06
Oct 1992 58.86 -484.78 365.30 -4.71 -321.09
Nov 1992 -506.07 355.36 -11.56 -327.84 -49.76
Dec 1992 551.08 103.76 -245.20 24.53 -155.19
Jan 1993 -122.62 -366.76 -68.22 -206.03 133.59
Feb 1993 -315.66 -36.65 -199.35 145.18 1325.53
Mar 1993 86.16 -156.19 180.29 1346.87 2007.63
Apr 1993 -187.18 165.14 1335.80 1998.04 514.30
May 1993 233.09 1376.64 2026.84 535.51 NA
Jun 1993 1290.56 1983.82 506.36 NA NA
Jul 1993 1500.17 277.60 NA NA NA
Aug 1993 -351.36 NA NA NA NA
Sep 1993 NA NA NA NA NA
```

```
CV_arima_boxcox %>% round(2) %>% tail(20)
```

```
h=1 h=2 h=3 h=4 h=5
Feb 1992 NA NA NA NA NA
Mar 1992 NA NA NA NA NA
Apr 1992 NA NA NA NA NA
May 1992 NA NA NA NA NA
Jun 1992 NA NA NA NA NA
Jul 1992 NA NA NA NA NA
Aug 1992 29.47 631.54 73.31 -685.90 344.30
Sep 1992 604.06 293.74 -540.11 125.87 -93.52
Oct 1992 -22.37 -404.99 272.42 -219.79 -238.68
Nov 1992 -352.38 307.37 -183.97 -290.77 159.81
Dec 1992 44.06 -159.34 -4.10 327.56 68.05
Jan 1993 -192.78 -23.95 303.46 68.61 85.15
Feb 1993 NA NA NA NA NA
Mar 1993 343.94 150.93 161.71 1245.97 1880.07
Apr 1993 89.52 84.71 1080.16 1737.25 351.10
May 1993 154.19 1592.55 2005.80 359.94 NA
Jun 1993 1043.97 1685.67 300.93 NA NA
Jul 1993 1525.71 186.20 NA NA NA
Aug 1993 -68.28 NA NA NA NA
Sep 1993 NA NA NA NA NA
```

```
CV_ets %>% round(2) %>% tail(20)
```

```
h=1 h=2 h=3 h=4 h=5
Feb 1992 NA NA NA NA NA
Mar 1992 NA NA NA NA NA
Apr 1992 NA NA NA NA NA
May 1992 NA NA NA NA NA
Jun 1992 NA NA NA NA NA
Jul 1992 NA NA NA NA NA
```

```
Aug 1992 NA NA NA NA NA
Sep 1992 NA NA NA NA NA
Oct 1992 NA NA NA NA NA
Nov 1992 NA NA NA NA NA
Dec 1992 NA NA NA NA NA
Jan 1993 NA NA NA NA NA
Feb 1993 NA NA NA NA NA
Mar 1993 NA NA NA NA NA
Apr 1993 NA NA NA NA NA
May 1993 NA NA NA NA NA
Jun 1993 NA NA NA NA NA
Jul 1993 NA NA NA NA NA
Aug 1993 NA NA NA NA NA
Sep 1993 NA NA NA NA NA
```

```
CV_ets_boxcox %>% round(2) %>% tail(20)
```

```
h=1 h=2 h=3 h=4 h=5
Feb 1992 NA NA NA NA NA
Mar 1992 NA NA NA NA NA
Apr 1992 NA NA NA NA NA
May 1992 NA NA NA NA NA
Jun 1992 NA NA NA NA NA
Jul 1992 NA NA NA NA NA
Aug 1992 NA NA NA NA NA
Sep 1992 NA NA NA NA NA
Oct 1992 NA NA NA NA NA
Nov 1992 NA NA NA NA NA
Dec 1992 NA NA NA NA NA
Jan 1993 NA NA NA NA NA
Feb 1993 NA NA NA NA NA
Mar 1993 NA NA NA NA NA
Apr 1993 NA NA NA NA NA
May 1993 NA NA NA NA NA
Jun 1993 NA NA NA NA NA
Jul 1993 NA NA NA NA NA
Aug 1993 NA NA NA NA NA
Sep 1993 NA NA NA NA NA
```

## Resultados

```
h = 5
#comparar os 4 modelos com os modelos abaixo

fit_auto_arima = auto.arima(serie) %>% forecast() %>% accuracy()
fit_ses = ses(serie, h=h) %>% forecast() %>% accuracy()
fit_holt = holt(serie, h=h) %>% forecast() %>% accuracy()
fit_ets = ets(serie) %>% forecast() %>% accuracy()
fit_stlf = stlf(serie, h=h) %>% forecast() %>% accuracy()
fit_bats = bats(serie) %>% forecast() %>% accuracy()
fit_tbats = tbats(serie) %>% forecast() %>% accuracy()
```

```

acuracia <- rbind(fit_auto_arima[3],
 fit_ses[3],
 fit_holt[3],
 fit_ets[3],
 fit_stlf[3],
 fit_bats[3],
 fit_tbats[3])

modelos = c("auto arima","ses","holt","ets","stlf","bats","tbats")

resultados = cbind(modelos, round(acuracia,3))

kable(resultados,col.names = c("Modelo", "Acuracia"),align = "c")

```

Modelo	Acuracia
auto arima	490.566
ses	736.233
holt	737.822
ets	487.182
stlf	470.926
bats	458.351
tbats	469.631