

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.

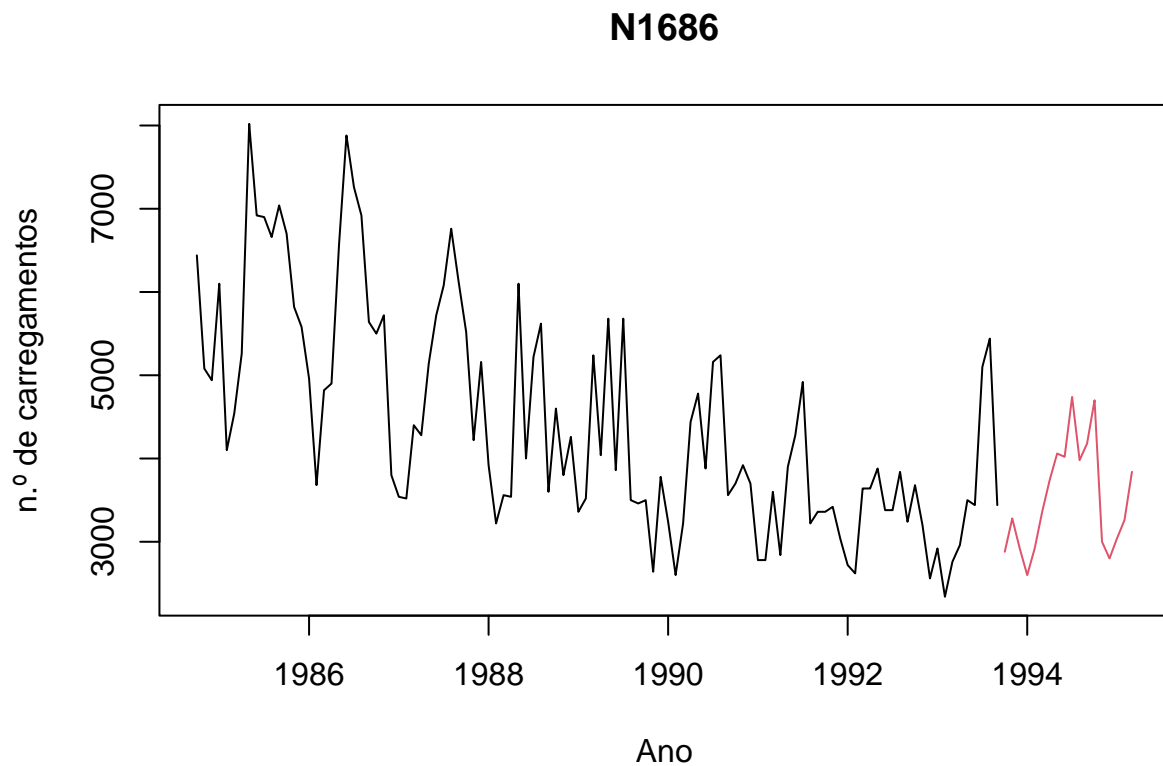


Figure 1: Comportamento da série ao longo do tempo

Decomposição MSTL

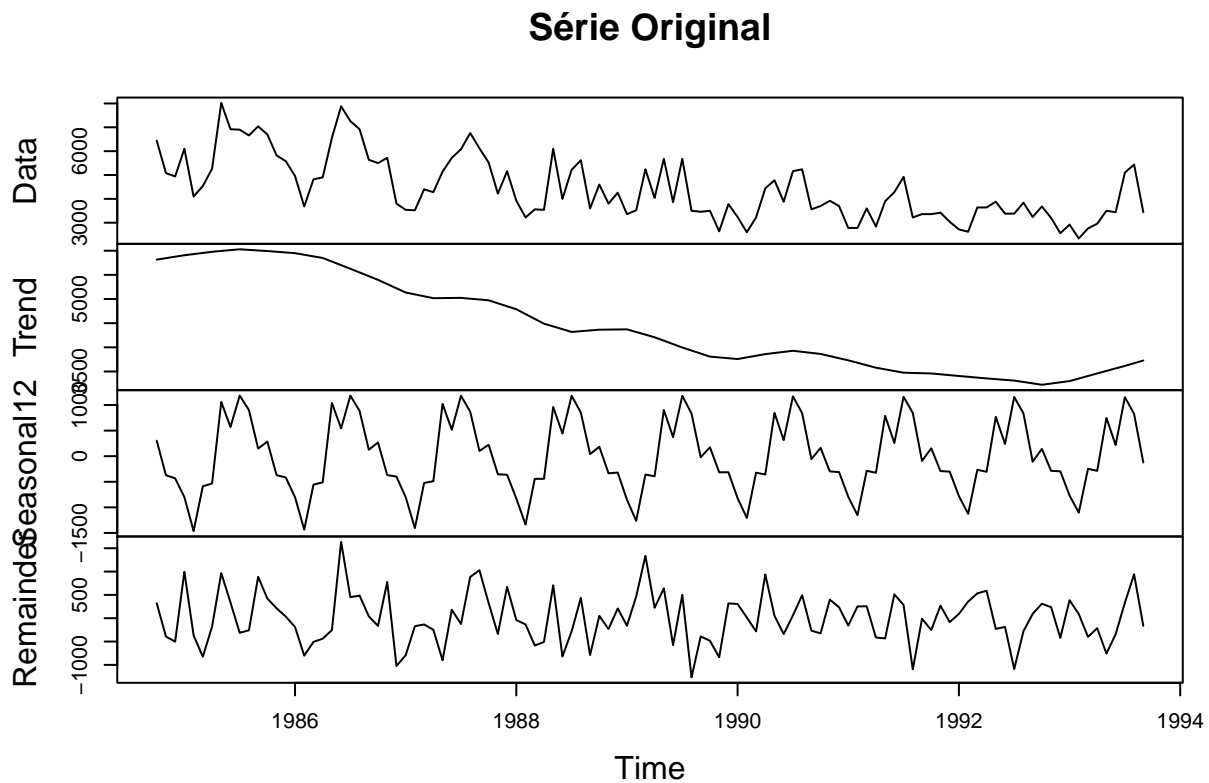


Figure 2: Decomposição MSTL

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.

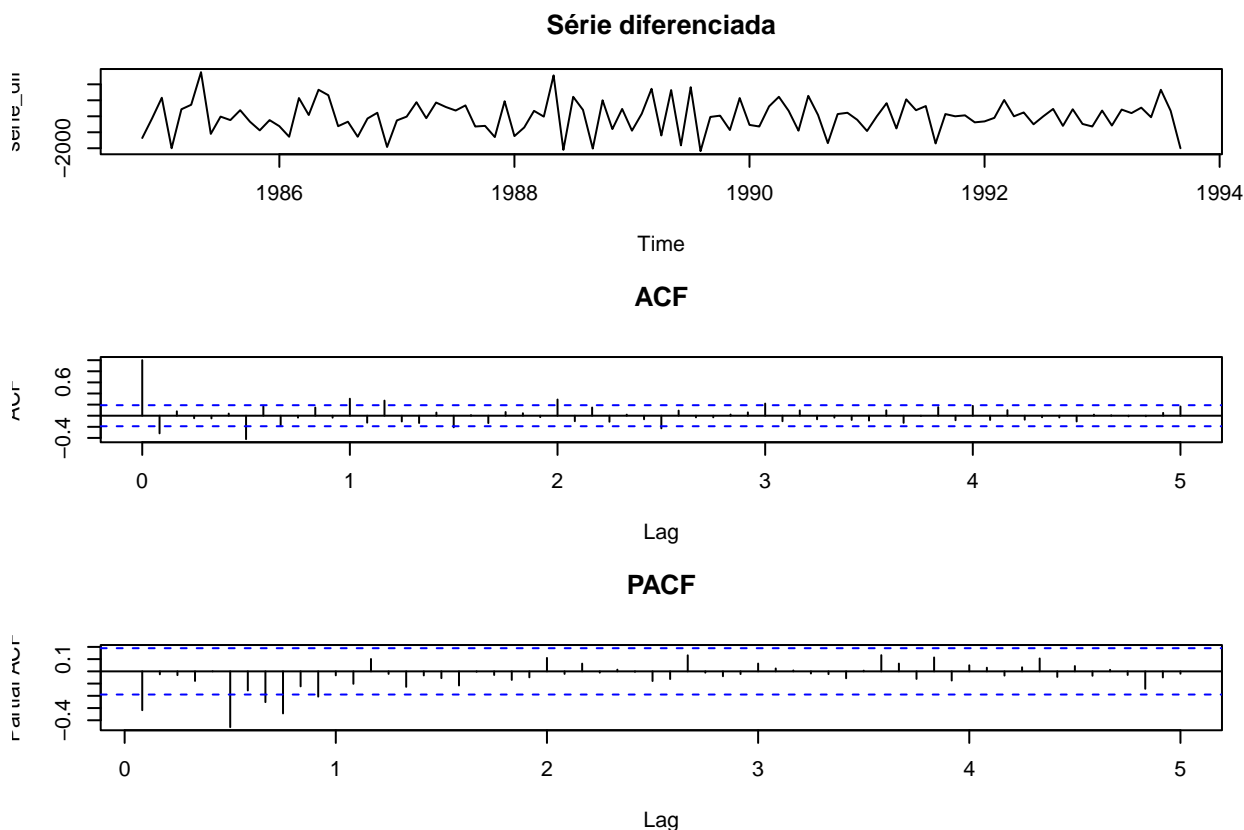


Figure 3: Gráficos ACF e PACF

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.

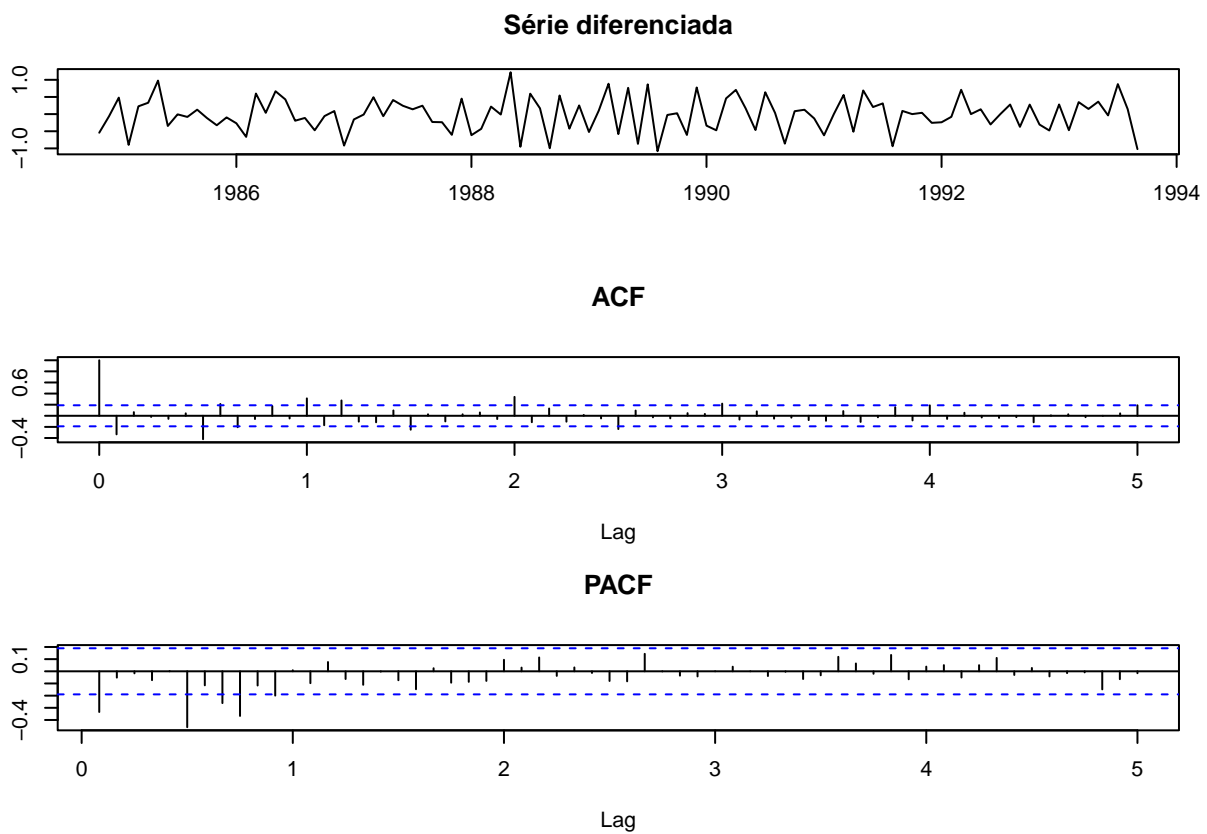


Figure 4: Gráficos ACF e PACF da série transformada (BoxCox)

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:

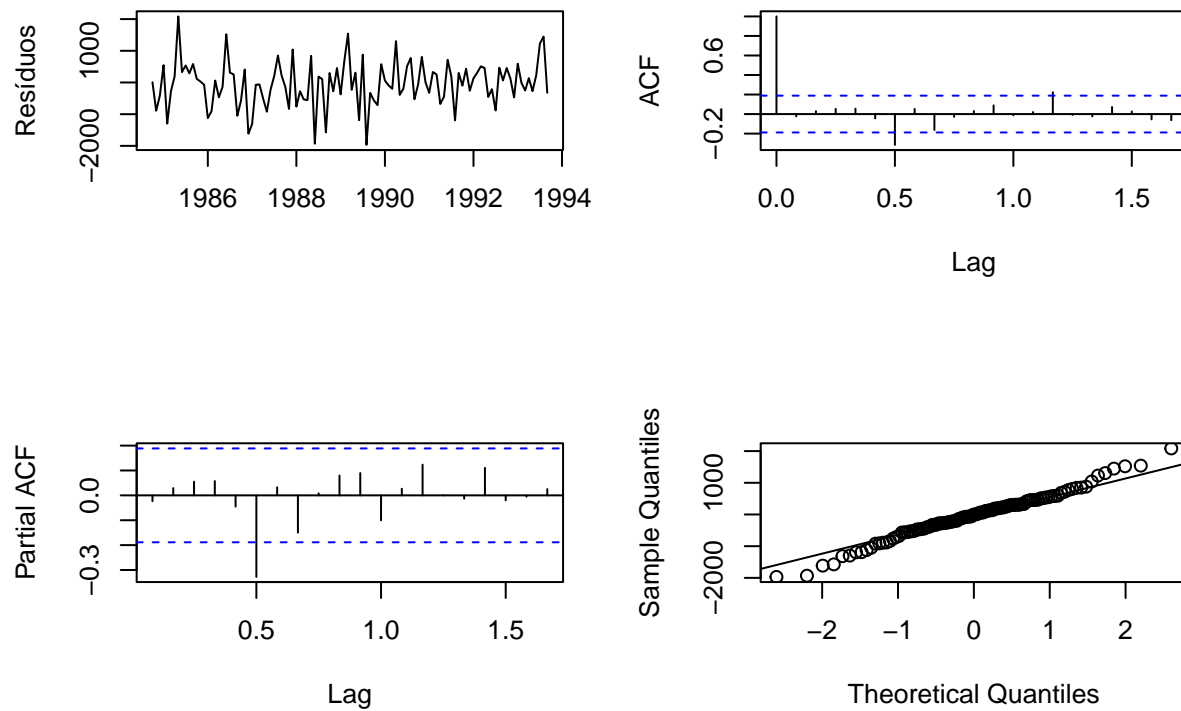


Figure 5: Resíduos ARIMA sem transformação

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

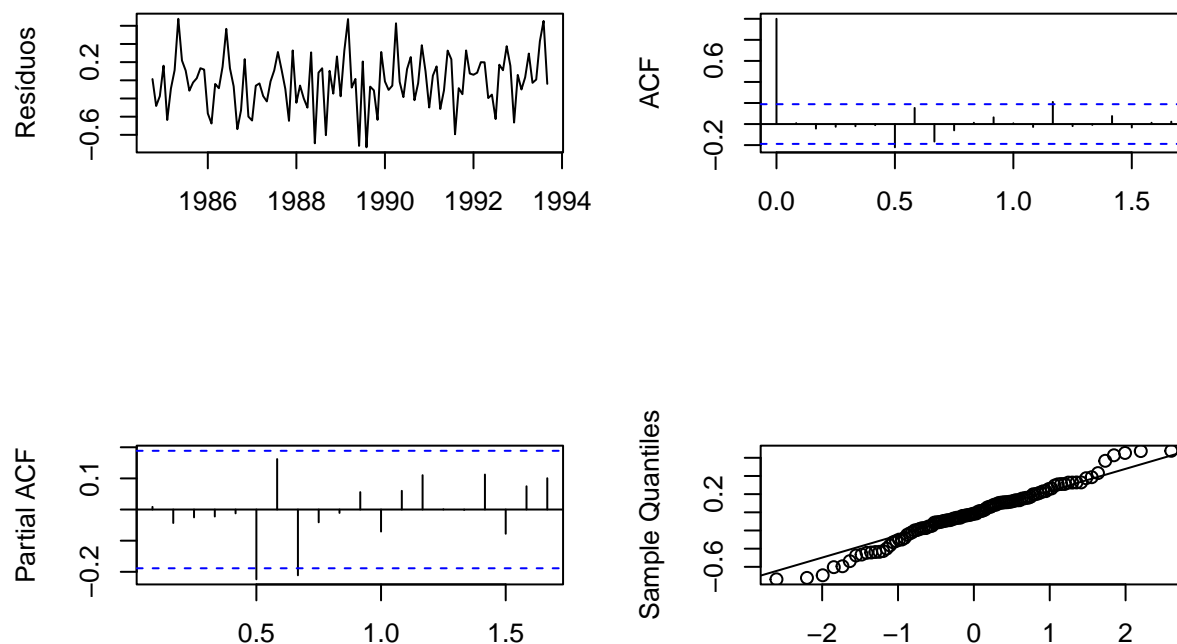


Figure 6: Resíduos ARIMA com transformação boxcox

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%.

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:

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

A decomposição do modelo ETS “MMM” e do modelo ETS “MAM com Dumped” e transformação de Box-Cox são ilustrados abaixo.

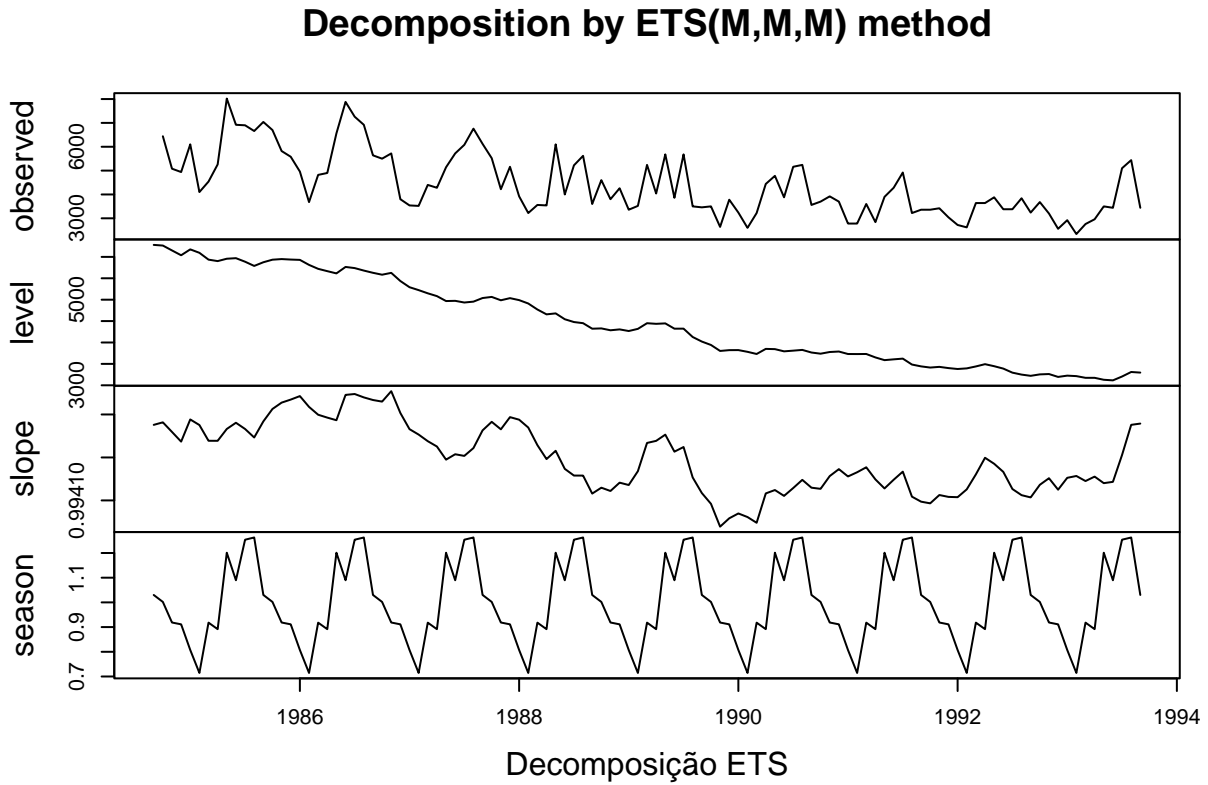


Figure 7: Decomposição ETS

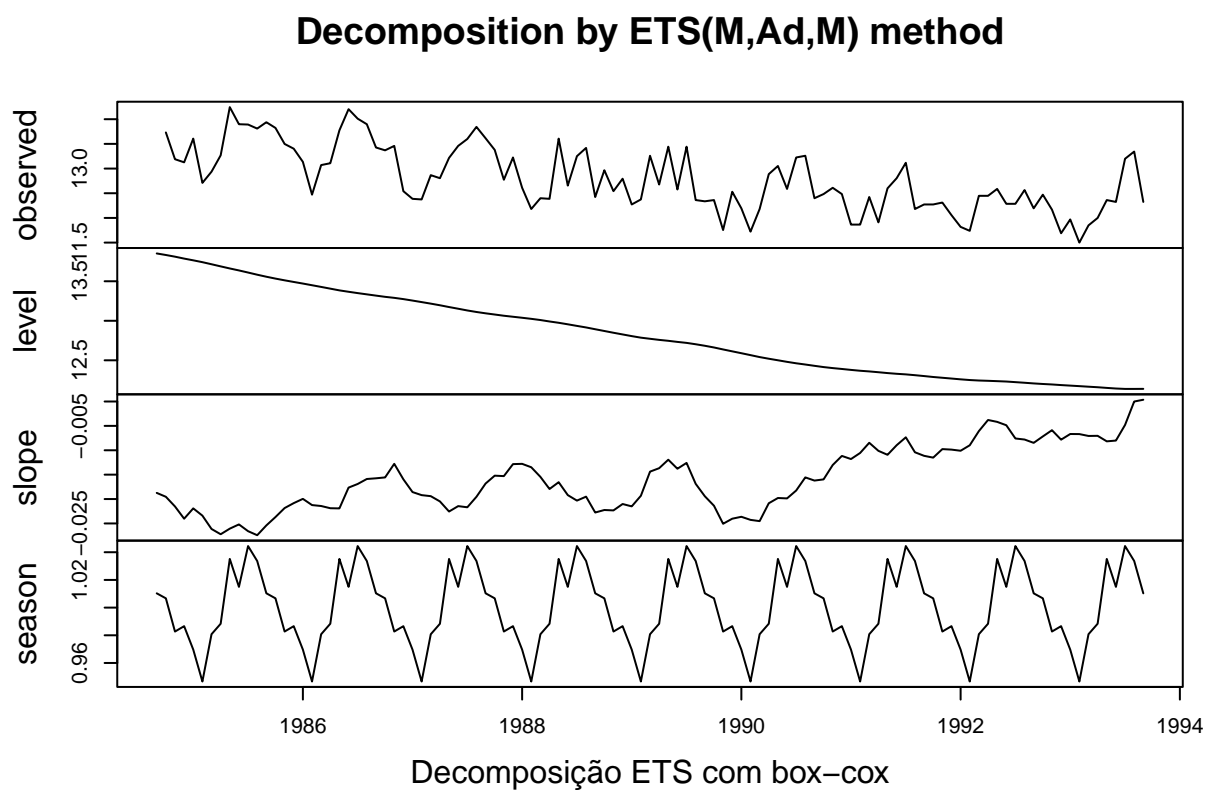


Figure 8: Decomposição ETS

Resíduos

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

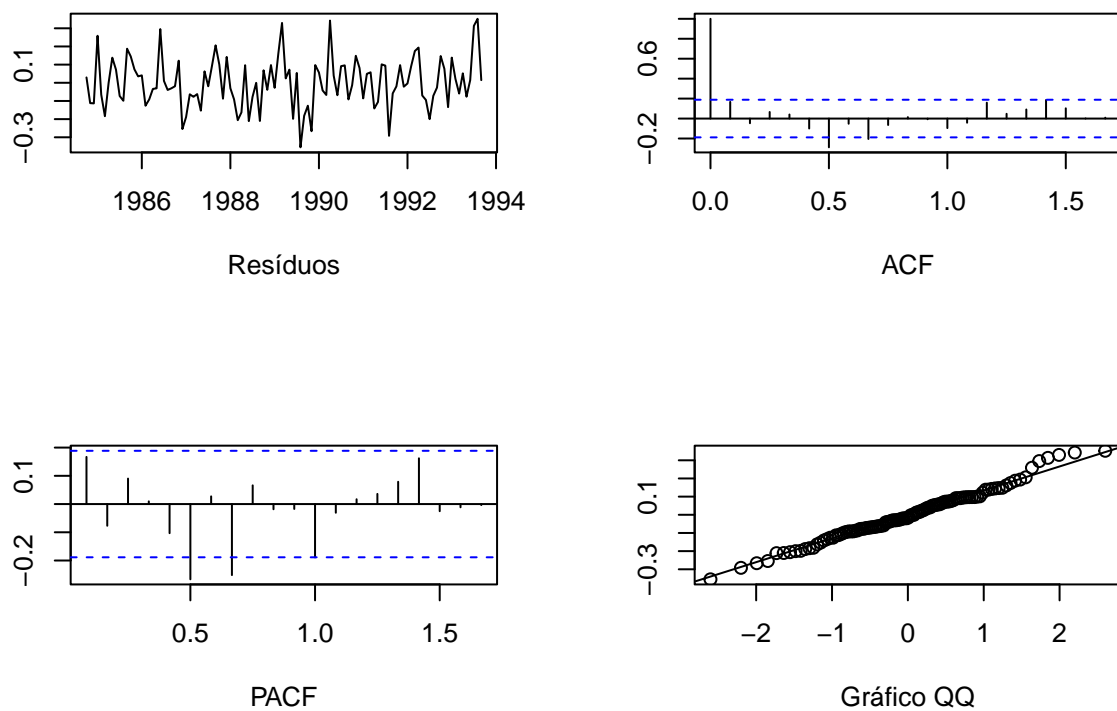


Figure 9: Análise de Resíduos do ETS

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

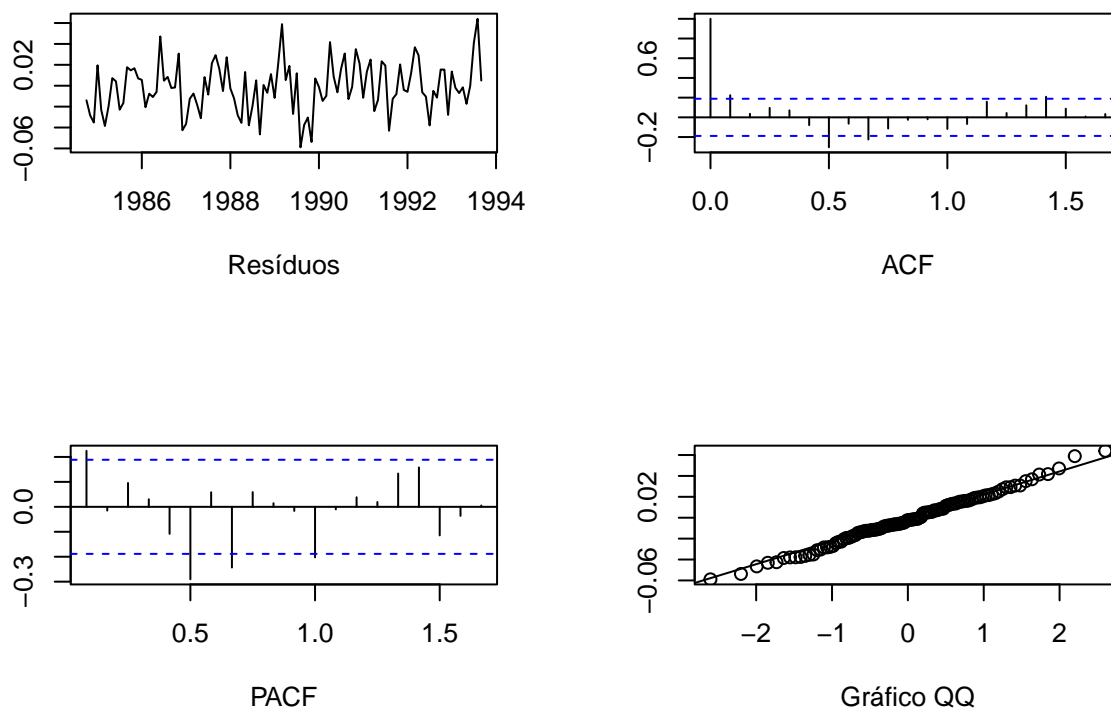


Figure 10: Análise de Resíduos do ETS do modelo com Box-Cox

Observa-se que os resíduos parecem comportar-se de maneira aleatória, com distribuição normal e sem autocorrelações importantes entre diferentes intervalos. Os testes formais encontram-se na tabela abaixo, a exemplo do anteriormente realizado:

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

Os resultados acima apresentados sugerem que os resíduos do modelo são estacionários, apresentam distribuição normal, entretanto, é possível a rejeição da hipótese nula de independência em ambos os casos.

Estudo de desempenho preditivo por janela deslizante

Considerante os dois modelos ARIMA e os dois modelos ETS anteriormente escolhidos, será realizado um estudo de desempenho preditivo por janela deslizante para a série M3 com ID 1686 com conjunto de treino consistindo nos últimos 14 períodos e considerando um horizonte preditivo de até 5 meses a frente. Para tal será utilizado a função `tsCV`, que calcula os erros de previsão obtidos aplicando a função de previsão a subconjuntos da série temporal, conforme ilustrado na figura abaixo. Os modelos utilizados para previsão são aqueles apresentados anteriormente: ARIMA, ARIMA com Box-Cox, ETS e ETS com box-cox.

	1	2	3	n-12	n-13	n-14						...	n
Passo 1										1	2	3	4	5		
Passo 2											1	2	3	4	5	
Passo 3												1	2	3	4	5
...													1	2	3	4
...														1	2	3
...															1	2
...																1

Figure 11: Janelas deslizantes

Para comparação dos resultados, foi calculado o erro absoluto médio (MAE) para horizonte de previsão.

Table 6: Erro absoluto médio

	ARIMA	ARIMA Box-Cox	ETS	ETS Box-Cox
h=1	618.7394	650.2503	607.9737	607.6476
h=2	682.1576	655.0000	689.5125	735.5779
h=3	721.7444	676.1780	694.9449	701.0805
h=4	897.2101	852.5248	906.0159	869.5573
h=5	1269.6113	1336.3204	1257.1669	1213.1535

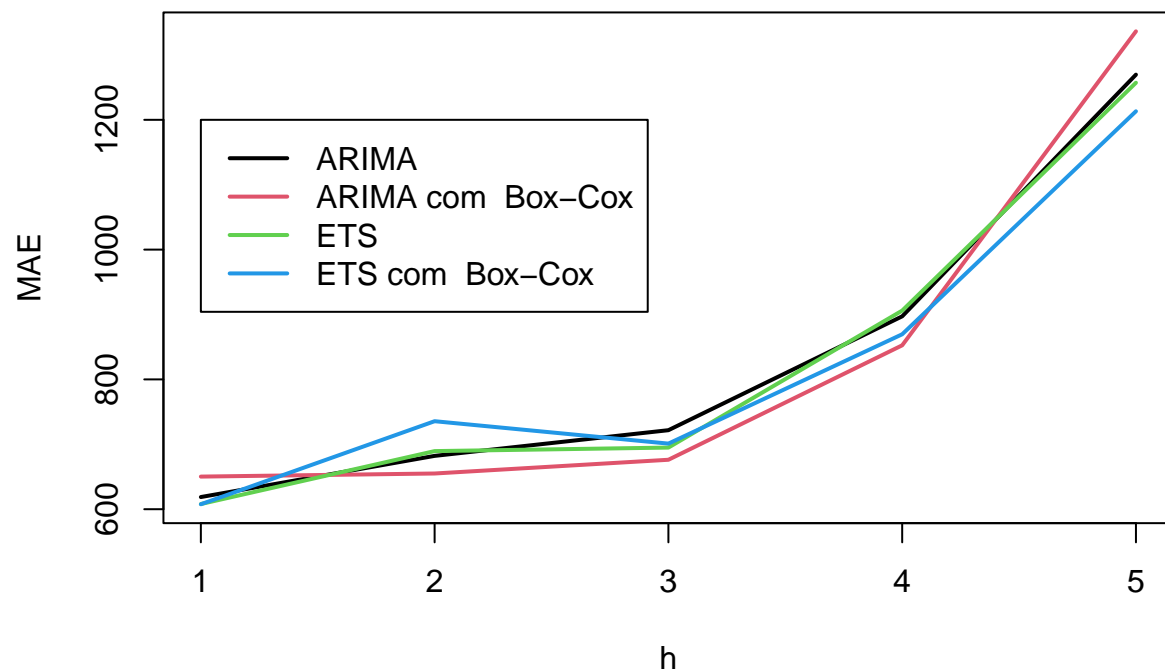
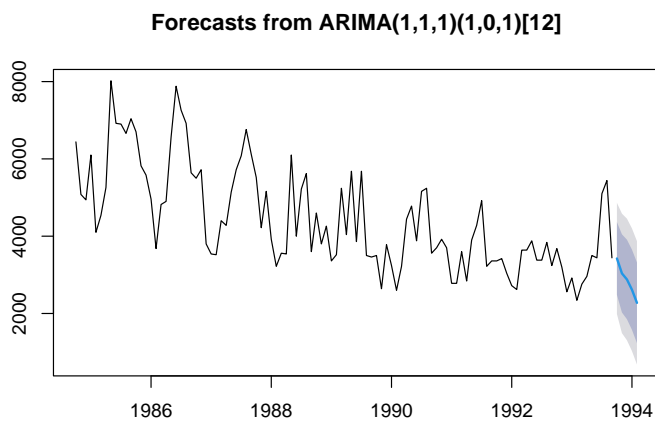


Figure 12: Erros absolutos médios a cada horizonte de previsão

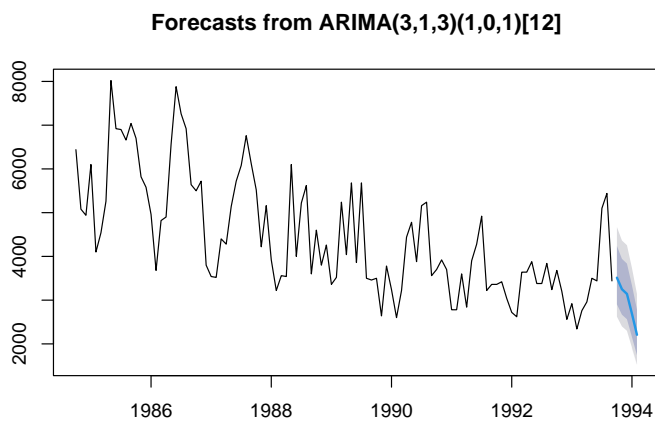
Resultados

Gráfico com as previsões de cada modelo

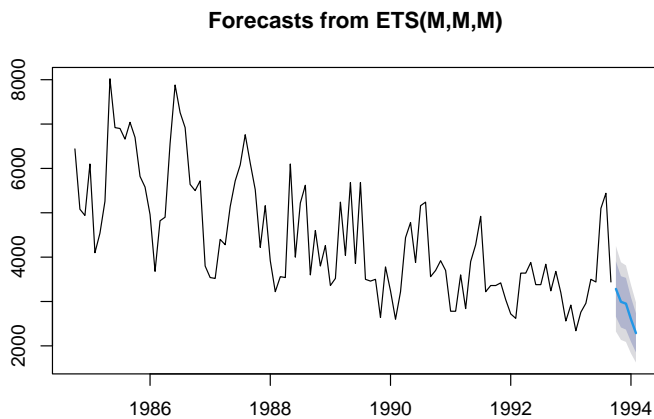
```
h = 5  
plot(f_arima(serie,h));plot(f_arima_boxcox(serie,h));
```



```
## Warning in InvBoxCox(pred$pred, lambda, biasadj, pred$se^2): biasadj
## information not found, defaulting to FALSE.
```



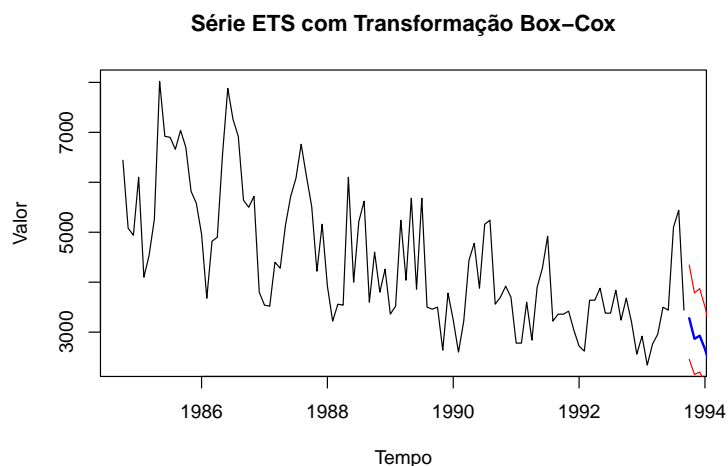
```
plot(f_ets(serie,h));
```



```
# fit_ets_boxcox <- ets(BoxCox(serie,lambda=lambda_boxcox), model = "MAM", damped = T)
previsao = forecast(fit_ets_boxcox,5,lambda=lambda_boxcox)
```

```
## Warning in forecast.ets(fit_ets_boxcox, 5, lambda = lambda_boxcox): biasadj
## information not found, defaulting to FALSE.
```

```
# Gráfico da série original e previsão
plot(serie, main = "Série ETS com Transformação Box-Cox",
     ylab = "Valor", xlab = "Tempo")
lines(previsao$mean, col = "blue", lwd = 2) # Linha da previsão
lines(previsao$upper[,2], col = "red") # Linha do intervalo superior
lines(previsao$lower[,2], col = "red")
```



Comparando os 4 modelos estimados anteriormente com os seguintes modelos: Auto Arima, SES, Holt, Auto ETS, STL, Bats, Tbat, temos os seguintes da métrica MAE, no conjunto de testes:

Table 7: Comparação do Erro absoluto médio com todos os modelos

Modelos	MAE
ARIMA	297.242
ARIMA com Box-Cox	337.326
ETS	271.07
ETS com Box-Cox	294.545
Auto ARIMA	692.388
SES	1096.291
Holt	1051.287
Auto ETS	444.531
STLF	429.104
Bats	397.695
Tbats	418.812

O modelo ETS(M,M,M) foi o que obteve o menor MAE no conjunto de teste, apresentando um valor igual a 271.07.