

Trabalho Prático 2

Análise de Séries Temporais - 1/2023

Ana Carolina Vianna - 18/0097261

César Augusto Galvão - 19/0011572

Yan Flávio Vianna - 14/0166149

Table of contents

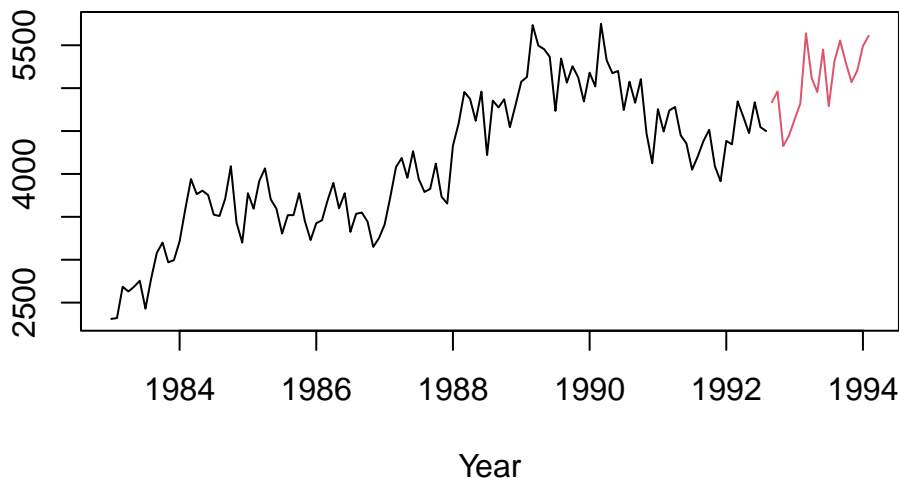
Introdução: série selecionada, características e decomposição	2
Modelos ARIMA: seleção, transformações e resíduos	3
Modelo sem transformação	3
Modelo com transformação	5
Modelos ETS: seleção, transformações e resíduos	7
Modelo sem transformação	7
Modelo com transformação	10
Estudo de desempenho preditivo	11
Resultados da Janela Deslizante	13
Performance em relação aos horizontes de previsão	13
Gráficos da previsão pontual e da previsão intervalar dos 4 modelos selecionados	14
Resultados	18
Apêndice	18

Introdução: série selecionada, características e decomposição

A série temporal escolhida foi a de número *id* correspondente a 2183. De acordo com a definição do próprio pacote, refere-se a *Fluid power shipments - hydraulic index*. Foram realizadas medidas mensais de 1983 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.

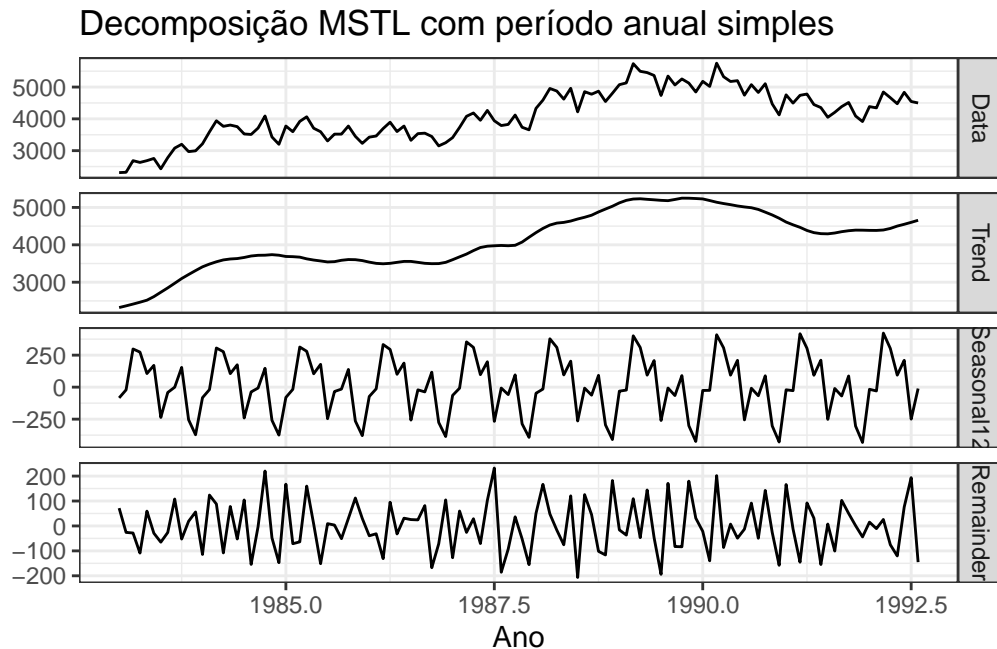
Série Temporal M3-2183



A série aparenta ter dois períodos, pelo menos: um ciclo anual e outro que compreende um período maior. No entanto, ao se tentar decompor a série com múltiplas sazonalidades, obté-se o seguinte:

- **Adicionando uma componente sazonal com ciclo menor que 1 ano** – uma das componentes sazonais apresenta heteroscedasticidade;
- **Adicionando uma componente sazonal com ciclo maior que 1 ano** – resíduos apresentam periodicidade ou heteroscedasticidade.

Optou-se portanto pela decomposição STL (apesar de os dados terem inicialmente formado um objeto `msts`) apenas com a sazonalidade anual, mas fica evidente que esta decomposição não é adequada quando se avalia a componente de tendência, que aparenta ainda carregar algum componente periódico. Os resíduos aparentam um comportamento aleatório e têm média - 0.104, o que é próximo de zero o suficiente considerando a magnitude dos dados da série. A decomposição é exposta a seguir.



Modelos ARIMA: seleção, transformações e resíduos

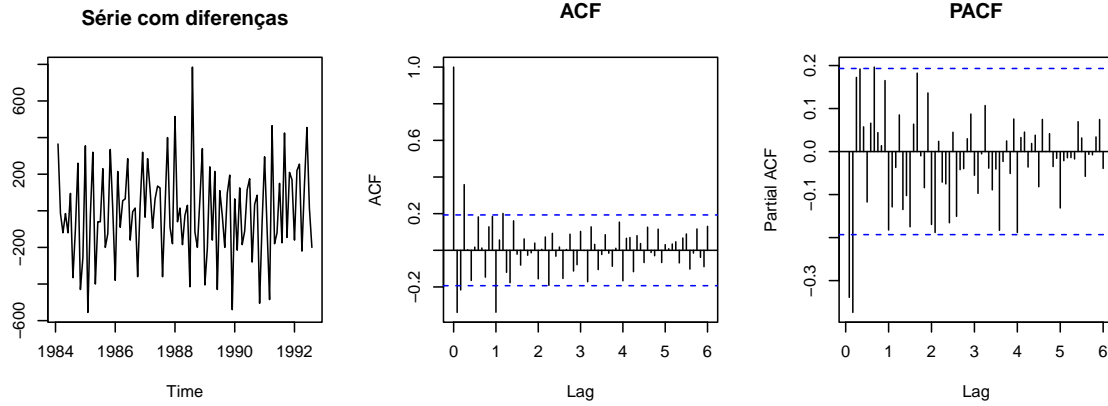
Modelo sem transformação

Seleção

Primeiramente, utilizou-se as funções *ndiffs()* e *nsdiffs()* do pacote *forecast* para identificar quantas diferenças simples e sazonais seriam necessárias para que a série se tornasse estacionária. Concluiu-se pelo resultado dessas funções que são necessárias uma diferenciação simples e uma sazonal. O teste KPSS confirma isso ao não rejeitar a hipótese nula de estacionariedade da série (com diferenças já aplicadas) ao nível de 5% de significância.

	Estatística	p-valor
KPSS Test for Level Stationarity	0.11	0.1

Prosseguimos com a seleção do melhor modelo ARIMA avaliando os gráficos de ACF e PACF. O primeiro parece apresentar quebra no primeiro lag sazonal, enquanto o segundo tem quebra no segundo lag simples, configurando um $ARIMA(2, 1, 0) \times (0, 1, 1)_{12}$ (porém os resíduos para este modelo não ficam muito bons). Entretanto, como não fica nítido um comportamento de queda amortizada, preferiu-se utilizar outro critério para a seleção do modelo.

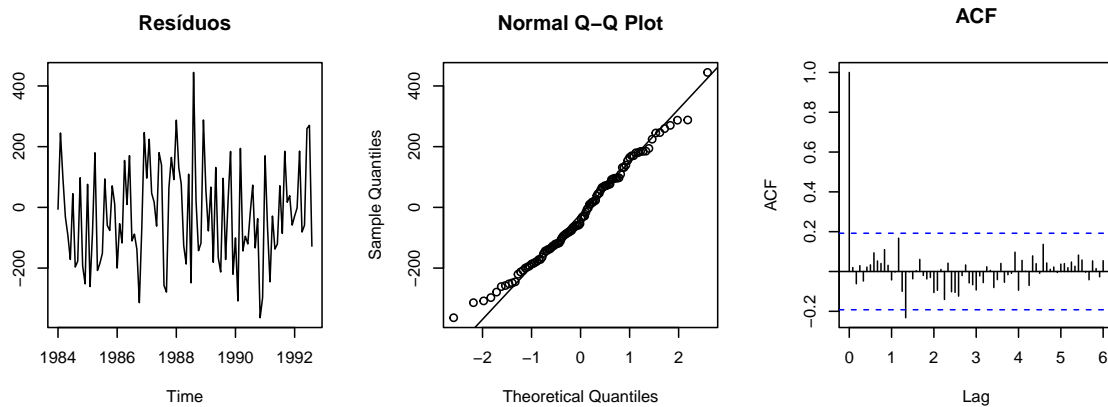


Optou-se pela varredura de combinações de p , q , P e Q , com d e D fixados em 1, como resultado das diferenciações já avaliadas. Utilizando o critério de Akaike corrigido, seleciona-se o modelo $\text{ARIMA}(2, 1, 2) \times (0, 1, 2)_{12}$ para a série, que possui o menor escore entre os modelos testados.

Ao se utilizar a função `auto.arima()`, recebe-se um modelo sugerido $\text{ARIMA}(2, 1, 2) \times (2, 1, 0)_{12}$, porém com AICc superior àquele identificado na varredura. Opta-se pelo modelo selecionado manualmente.

Resíduos

Foram retirados os zeros da inicialização para possibilitar a análise dos resíduos. Observa-se pelo gráfico que os resíduos são aleatórios e aparentemente centrados em zero, com variação constante. Além disso, verifica-se uma distribuição aproximadamente normal, mas com caudas mais pesadas. Finalmente, o gráfico ACF apresenta que a autocorrelação dos resíduos está, em sua grande maioria, dentro da banda de confiança, com exceção de um ponto, que extrapola ligeiramente a margem.



Por fim, realiza-se testes de hipótese para independência e normalidade (o teste KPSS para estacionariedade já foi apresentado) e seus resultados são apresentados na tabela a seguir. De fato, o teste de Shapiro-Wilk não rejeita a normalidade da distribuição dos resíduos apesar de o gráfico QQ apresentar caudas pesadas. Além disso, o teste Ljung-Box com *lag* igual a 15 também não rejeita a independência entre os resíduos e, conseqüentemente, os dados da série.

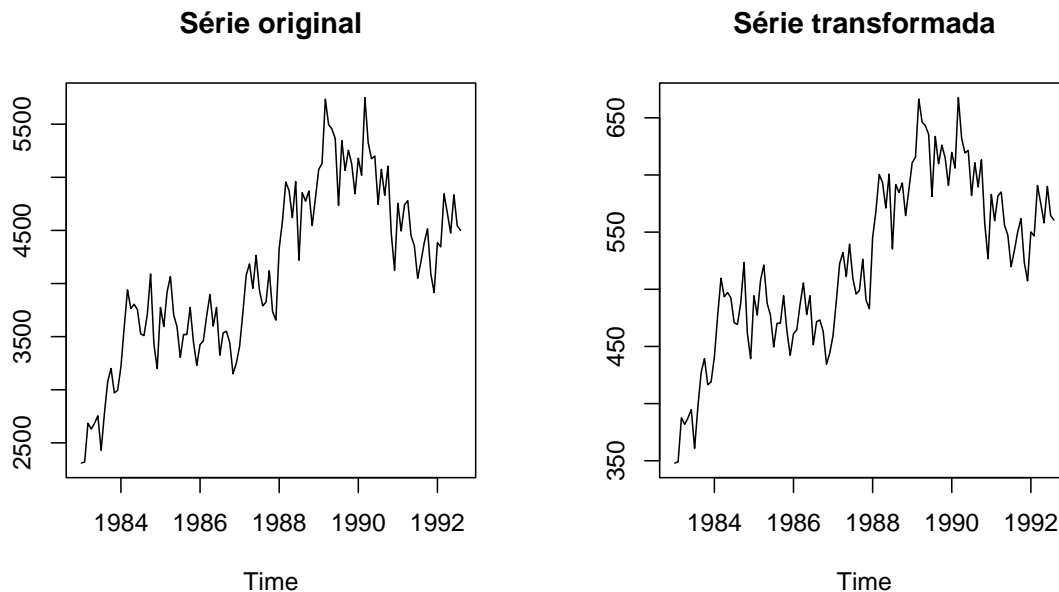
	Estatística	p-valor	Lag
Box-Ljung test	8.90	0.88	15
Shapiro-Wilk normality test	0.99	0.35	

Modelo com transformação

Seleção

Foi utilizada a função *BoxCox.lambda()* do pacote *forecast* para decidir de forma automatizada o melhor valor de lambda para a transformação de Box-Cox. A função sugere um valor de $\lambda = 0.71$.

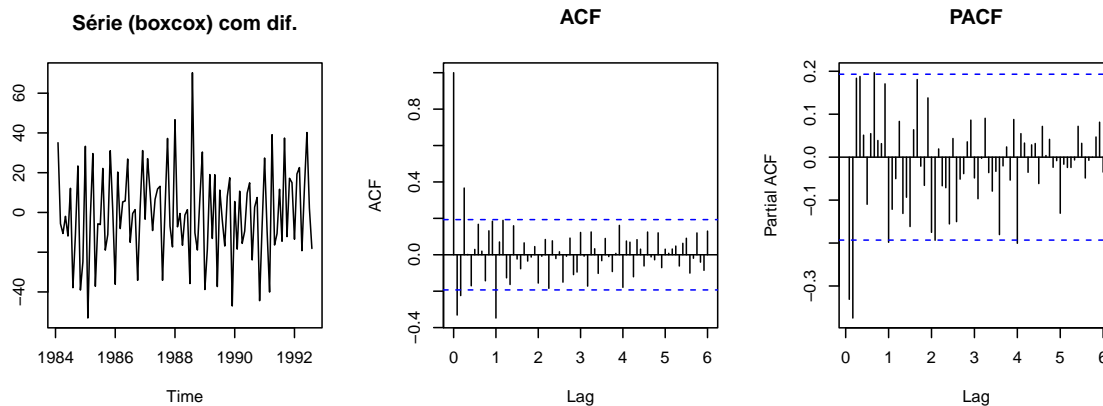
Apesar de haver uma sugestão de transformação, não é possível avaliar graficamente se houve uma diferença significativa no comportamento da série temporal excetuando-se a escala, como se pode ver nos eixos dos gráficos a seguir.



Após aplicar a transformação de Box-Cox na série, utilizou-se as funções `ndiffs()` e `nsdiffs()` para identificar quantas diferenciações simples e sazonais seriam necessárias para que a série se torne estacionária. Concluiu-se que são necessárias uma diferenciações simples e uma diferenciações sazonal, o que é confirmado pelo resultado do teste KPSS nos resíduos da série com as diferenças já aplicadas.

	Estatística	p-valor
KPSS Test for Level Stationarity	0.12	0.1

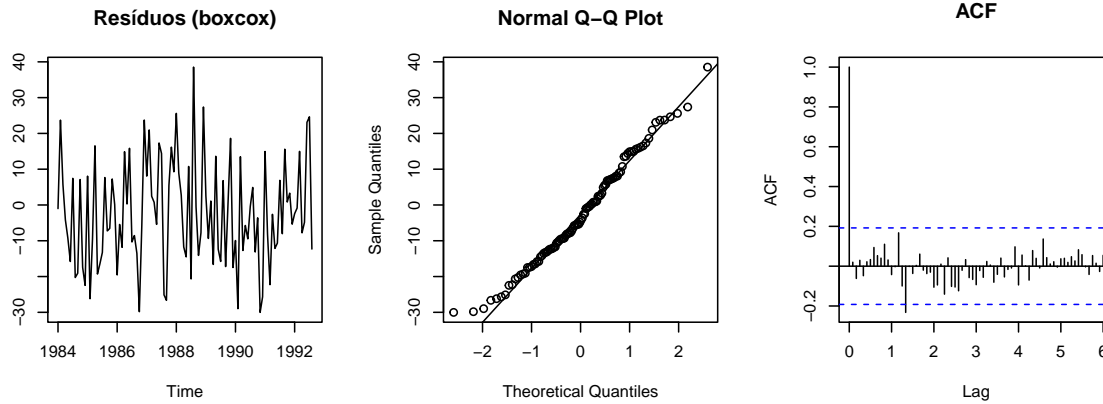
O gráfico da ACF parece apresentar quebra no primeiro lag sazonal, enquanto o PACF tem quebra no segundo lag simples, o que configura um $ARIMA(2, 1, 0) \times (0, 1, 1)_{12}$ (porém, mais uma vez, os resíduos para este modelo não ficam muito bons). Entretanto, os gráficos não evidenciam comportamentos claros para a série. Novamente, os resíduos parecem ter média igual a zero.



Foram testadas combinações de p , q , P e Q , com d e D fixados em 1 e, em seguida, selecionou-se o modelo ARIMA que apresentava menor valor do AICc. Temos, então, que o modelo escolhido para a série transformada é um $ARIMA(2, 1, 2) \times (0, 1, 2)_{12}$, assim como no caso da série sem transformação. Utilizando-se a função `auto.arima()` recebe-se uma sugestão de um modelo $ARIMA(3, 1, 1) \times (2, 1, 0)_{12}$ mas, assim como ocorre no modelo sem transformação, opta-se pelo modelo selecionado manualmente por apresentar um AICc menor.

Resíduos

Foram retirados os zeros da inicialização para seguir com a análise dos resíduos. O gráfico da série dos resíduos sugere aleatoriedade e o QQ plot distribuição aproximadamente normal. Por último, o gráfico ACF mostra que a autocorrelação dos resíduos está dentro da banda de confiança, com exceção de um ponto que excede um pouco este limite.



Assim como ocorre para a série não transformada, os testes de Shapiro-Wilk e Ljung-box com lag igual a 15 não apresentam indicação para rejeição de suas hipóteses nulas. Isto é, pode-se dizer que a série transformada tem distribuição normal e seus resíduos são independentes.

	Estatística	p-valor	Lag
Box-Ljung test	8.41	0.91	15
Shapiro-Wilk normality test	0.98	0.23	

Modelos ETS: seleção, transformações e resíduos

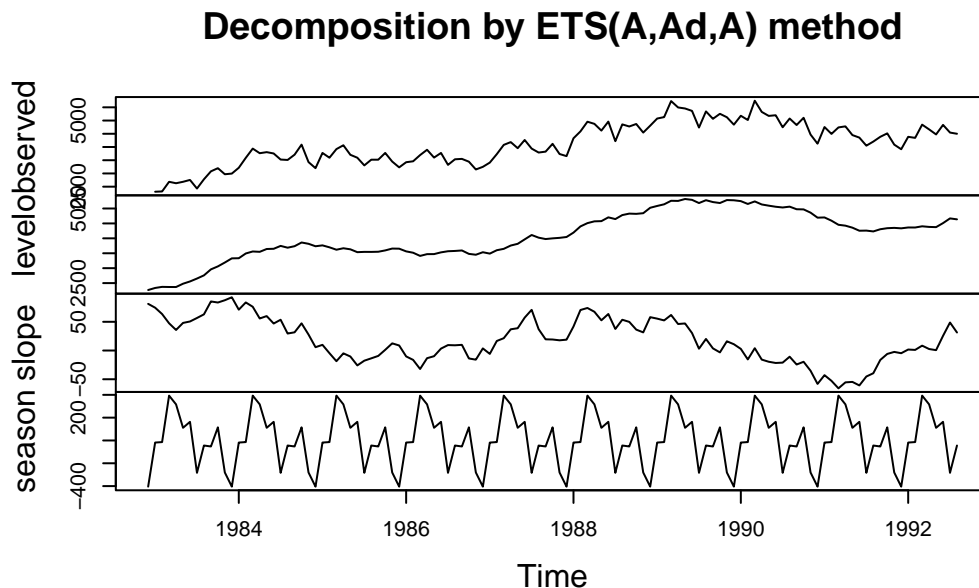
Modelo sem transformação

Seleção

Para a seleção do modelo ETS, foi realizada uma varredura com todas as combinações possíveis de erro, tendência e sazonalidade, assim como a aplicação ou não de *damp* na tendência. Os seis modelos com melhores indicadores são exibidos na tabela a seguir para comparação.

Modelo	AIC	AICc	BIC
ETS(M,Ad,M)	1761.30	1768.36	1810.87
ETS(M,M,M)	1761.94	1769.00	1811.51
ETS(A,Ad,A)	1764.25	1771.30	1813.81
ETS(M,Ad,A)	1767.73	1774.78	1817.29
ETS(M,A,M)	1769.04	1775.29	1815.86
ETS(A,A,A)	1771.20	1777.44	1818.01

De fato, modelos que misturam termos aditivos e multiplicativos apresentam os melhores indicadores AICc, mas o desempenho da função `ets()` é instável nessas condições. Por isso, opta-se pelo uso do terceiro melhor modelo, $ETS(A, Ad, A)$. A decomposição da série temporal analisada utilizando este modelo é exposta no gráfico a seguir.



Resíduos

Uma análise visual dos resíduos indica comportamento aleatório em torno de zero e autocorrelações próximas a zero. Quanto à distribuição, a amostra parece ter uma distribuição próxima à normal, mas com caudas mais pesadas.

Observa-se ainda pelo resultado dos testes de hipótese a seguir temos uma série estacionária, com resíduos normalmente distribuídos e mutuamente independentes.

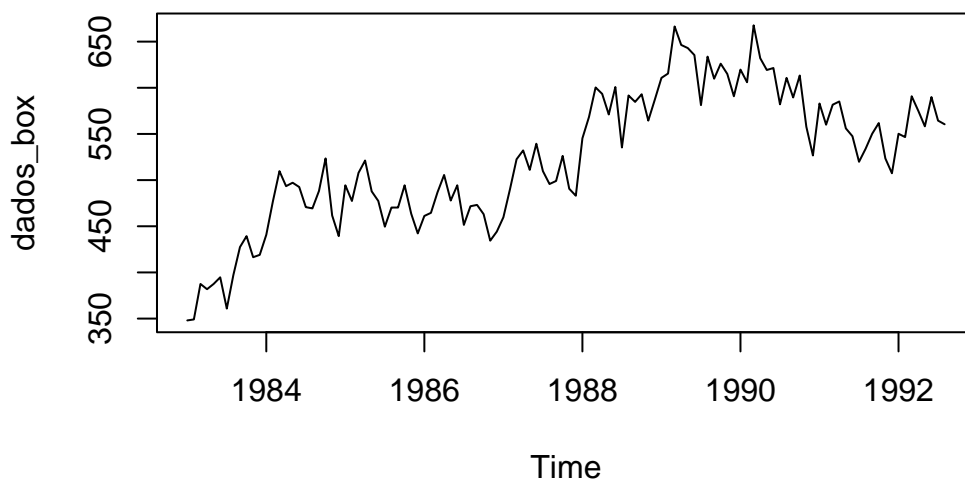
	Estatística	p-valor	Lag
KPSS Test for Level Stationarity	0.07	0.10	4
Box-Ljung test	21.35	0.13	15
Shapiro-Wilk normality test	0.99	0.38	

Modelo com transformação

Seleção

Utilizando a função `BoxCox.lambda()` obtém-se uma sugestão de transformação com $\lambda = 0.712$. A série transformada é exibida no gráfico a seguir.

Série com transformação Box–Cox $\lambda = 0.712$

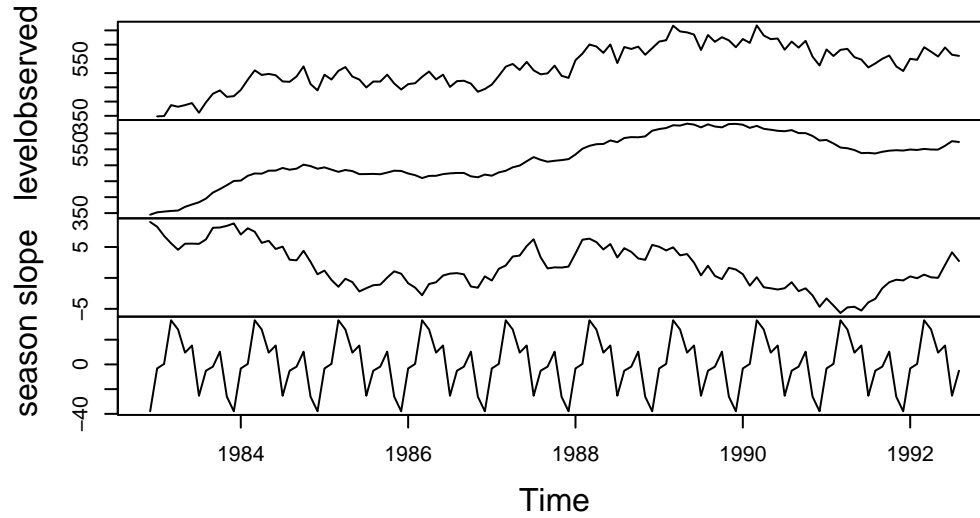


No caso do modelo com transformação BoxCox a recomendação de modelo, obtida pelo mesmo método de varredura, é primariamente a seleção já feita para o modelo sem transformação, que é o $ETS(A, Ad, A)$.

Modelo transformado	AIC	AICc	BIC
ETS(A,Ad,A)	1205.77	1212.82	1255.33
ETS(M,M,M)	1207.69	1214.75	1257.26
ETS(M,Ad,M)	1207.92	1214.98	1257.49
ETS(M,Ad,A)	1208.24	1215.29	1257.81
ETS(A,A,A)	1218.37	1224.62	1265.18
ETS(M,A,A)	1221.45	1227.69	1268.26

Os componentes do ajuste do modelo $ETS(A, Ad, A)$ são expostas a seguir e, como no caso anterior, a transformação não apresenta nenhuma mudança aparente no comportamento da série.

Decomposition by ETS(A,Ad,A) method



Resíduos

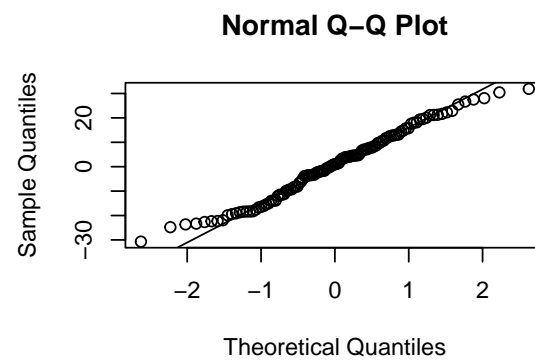
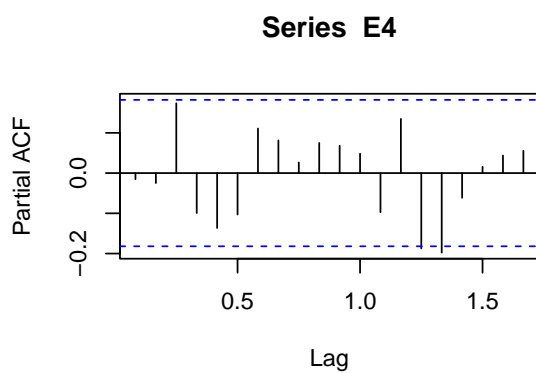
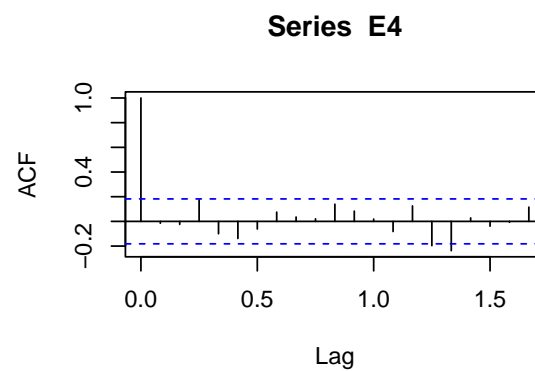
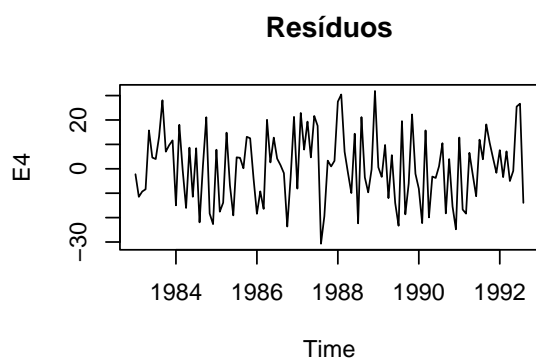
Assim como ocorre para a série não transformada, os resíduos parecem exibir comportamento aleatório em torno de zero, autocorrelações próximas a zero e distribuição próxima à normal, mas com caudas mais pesadas.

Novamente, os testes de hipótese expostos na tabela a seguir corroboram a hipótese de se tratar de uma série estacionária, com resíduos normalmente distribuídos e mutuamente independentes.

	Estatística	p-valor	Lag
KPSS Test for Level Stationarity	0.07	0.10	4
Box-Ljung test	20.48	0.15	15
Shapiro-Wilk normality test	0.98	0.16	

Estudo de desempenho preditivo

Para realizar a análise do desempenho preditivo usando uma abordagem de janela deslizante, o estudo considera uma janela de tamanho $n - 14$ e calcula os erros de previsão para horizontes de até 5 períodos. Utilizando os modelos previamente mencionados para criar as funções de



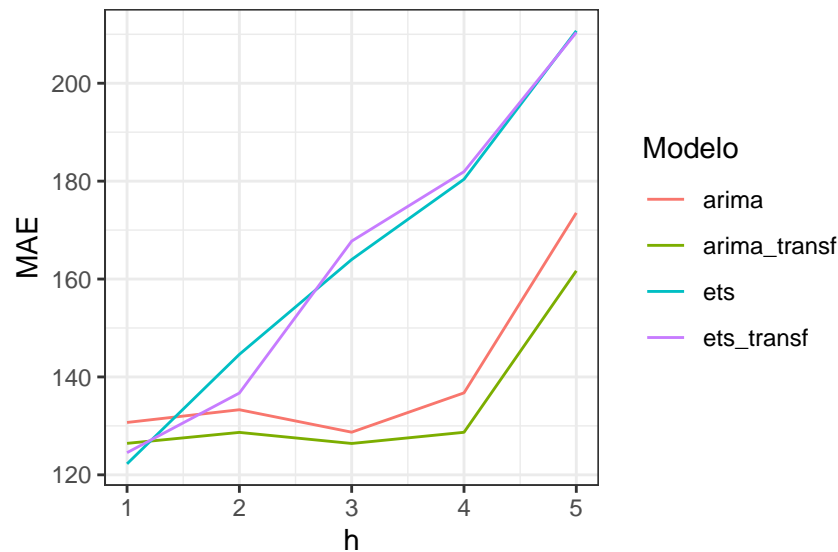
previsão, Os resultados são apresentados em um gráfico e uma tabela, mostrando os erros absolutos para cada horizonte de previsão.

Resultados da Janela Deslizante

	ARIMA	ETS	ARIMA Transformada	ETS Transformada
h=1	130.701	126.428	122.244	124.525
h=2	133.301	128.668	144.609	136.707
h=3	128.709	126.408	163.981	167.741
h=4	136.742	128.690	180.380	181.928
h=5	173.517	161.675	210.707	210.361

VER PQ A ETS_TRANSF ESTÁ DANDO TUDO NA

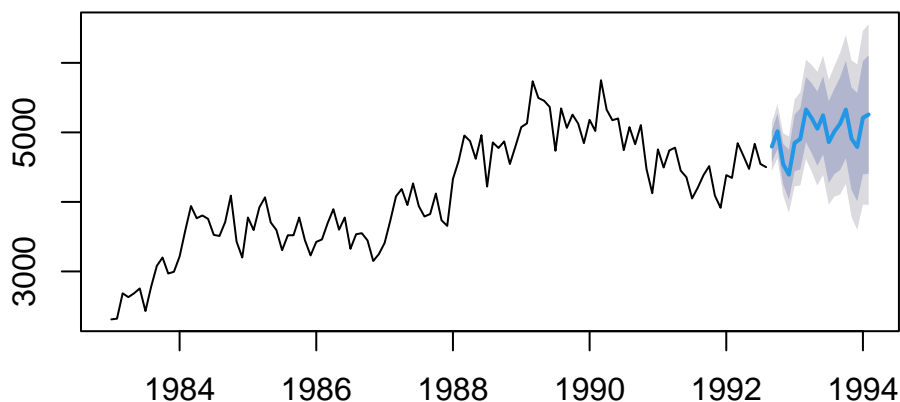
Performance em relação aos horizontes de previsão



ARRUMAR A LEGENDA E OS VALORES E DEPOIS FAZER A ANÁLISE

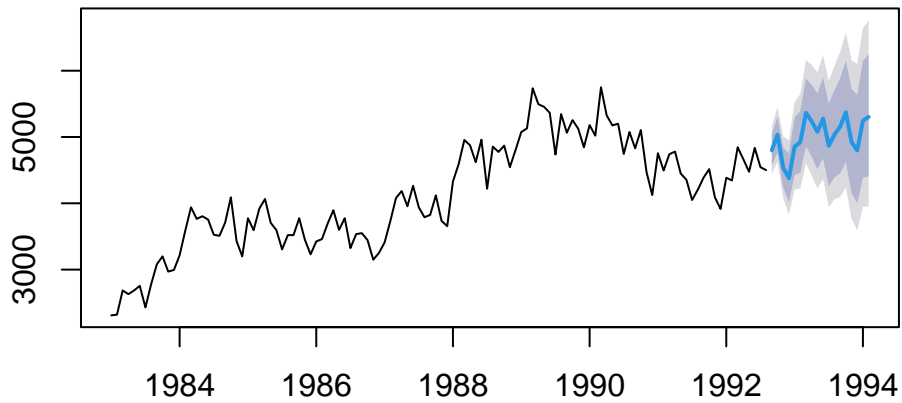
Gráficos da previsão pontual e da previsão intervalar dos 4 modelos selecionados

Previsão 95% de probabilidade de cobertura



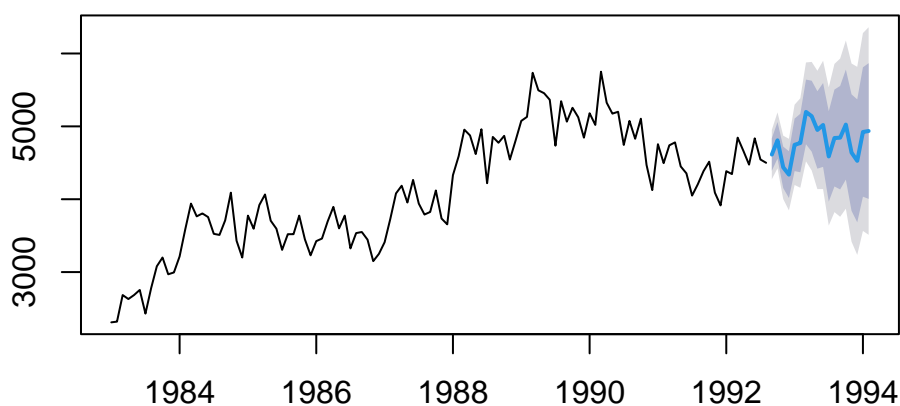
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Sep 1992	4795.360	4568.189	5022.531	4447.932	5142.788
Oct 1992	5016.382	4763.010	5269.753	4628.883	5403.880
Nov 1992	4541.278	4248.643	4833.913	4093.731	4988.825
Dec 1992	4391.441	4032.320	4750.562	3842.213	4940.669
Jan 1993	4848.159	4440.260	5256.059	4224.331	5471.988
Feb 1993	4905.299	4467.187	5343.411	4235.264	5575.333
Mar 1993	5329.729	4861.354	5798.103	4613.411	6046.046
Apr 1993	5203.385	4700.475	5706.295	4434.251	5972.520
May 1993	5052.665	4518.132	5587.198	4235.167	5870.163
Jun 1993	5245.419	4683.930	5806.909	4386.695	6104.144
Jul 1993	4859.079	4271.730	5446.428	3960.806	5757.352
Aug 1993	5014.812	4401.327	5628.297	4076.568	5953.057
Sep 1993	5125.554	4459.124	5791.984	4106.338	6144.771
Oct 1993	5328.565	4627.990	6029.140	4257.129	6400.001
Nov 1993	4908.795	4172.038	5645.551	3782.023	6035.566
Dec 1993	4788.711	4010.224	5567.197	3598.119	5979.303
Jan 1994	5210.324	4393.633	6027.015	3961.303	6459.345
Feb 1994	5256.719	4407.774	6105.663	3958.370	6555.067

Previsão 95% de probabilidade de cobertura



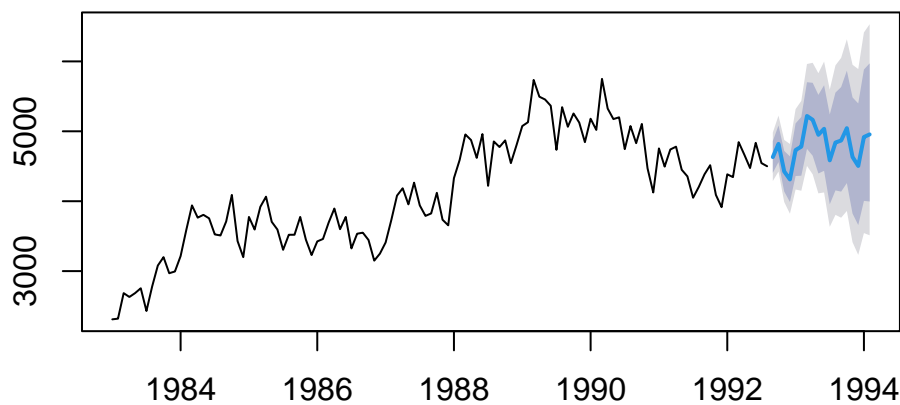
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Sep 1992	4799.752	4565.169	5037.686	4442.368	5164.973
Oct 1992	5035.236	4767.608	5307.025	4627.652	5452.555
Nov 1992	4537.145	4237.281	4842.830	4080.957	5006.952
Dec 1992	4379.433	4016.236	4751.524	3827.684	4951.992
Jan 1993	4854.191	4428.229	5291.205	4207.361	5526.884
Feb 1993	4922.088	4462.585	5394.299	4224.665	5649.249
Mar 1993	5364.435	4861.280	5881.571	4600.788	6160.802
Apr 1993	5235.766	4699.996	5787.823	4423.227	6086.423
May 1993	5077.223	4512.969	5660.159	4222.156	5976.018
Jun 1993	5275.453	4676.146	5895.059	4367.467	6230.952
Jul 1993	4867.736	4256.482	5501.964	3942.656	5846.601
Aug 1993	5035.622	4391.144	5704.808	4060.475	6068.613
Sep 1993	5155.398	4452.819	5886.738	4093.156	6284.981
Oct 1993	5372.068	4624.761	6150.631	4242.499	6574.823
Nov 1993	4923.688	4159.707	5723.503	3770.673	6160.632
Dec 1993	4797.716	3998.702	5637.111	3593.158	6096.874
Jan 1994	5246.596	4386.318	6149.642	3949.353	6644.026
Feb 1994	5302.608	4406.473	6244.738	3951.961	6761.014

Previsão 95% de probabilidade de cobertura



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Sep 1992	4613.023	4391.648	4834.399	4274.459	4951.588
Oct 1992	4808.078	4557.385	5058.771	4424.676	5191.480
Nov 1992	4437.827	4153.583	4722.072	4003.113	4872.541
Dec 1992	4337.828	4016.796	4658.859	3846.853	4828.803
Jan 1993	4746.470	4386.219	5106.721	4195.513	5297.426
Feb 1993	4771.433	4370.152	5172.715	4157.726	5385.140
Mar 1993	5198.160	4754.516	5641.804	4519.665	5876.655
Apr 1993	5139.633	4652.663	5626.602	4394.877	5884.388
May 1993	4950.914	4419.944	5481.884	4138.865	5762.962
Jun 1993	5019.380	4443.961	5594.799	4139.353	5899.408
Jul 1993	4587.096	3966.960	5207.232	3638.680	5535.512
Aug 1993	4840.204	4175.228	5505.179	3823.211	5857.196
Sep 1993	4844.615	4134.766	5554.463	3758.995	5930.234
Oct 1993	5024.388	4269.787	5778.989	3870.325	6178.451
Nov 1993	4639.864	3840.678	5439.050	3417.615	5862.114
Dec 1993	4526.533	3682.995	5370.072	3236.453	5816.614
Jan 1994	4922.724	4035.118	5810.330	3565.247	6280.201
Feb 1994	4936.057	4004.711	5867.404	3511.686	6360.429

Previsão 95% de probabilidade de cobertura



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Sep 1992	4631.450	4406.500	4859.593	4288.733	4981.634
Oct 1992	4819.800	4560.493	5083.189	4424.909	5224.240
Nov 1992	4426.133	4138.592	4719.160	3988.650	4876.449
Dec 1992	4311.700	3989.207	4641.298	3821.446	4818.576
Jan 1993	4732.398	4360.531	5112.881	4167.267	5317.689
Feb 1993	4780.196	4365.140	5205.907	4149.876	5435.446
Mar 1993	5219.170	4748.721	5702.169	4504.933	5962.773
Apr 1993	5164.246	4650.160	5693.522	4384.401	5979.640
May 1993	4949.949	4397.247	5521.041	4112.426	5830.514
Jun 1993	5035.976	4434.720	5658.674	4125.515	5996.633
Jul 1993	4584.535	3955.551	5239.447	3633.645	5596.136
Aug 1993	4839.699	4154.975	5553.574	3804.956	5942.701
Sep 1993	4868.346	4137.019	5632.826	3764.092	6050.245
Oct 1993	5044.941	4259.924	5866.888	3860.229	6316.151
Nov 1993	4632.659	3823.646	5484.664	3413.999	5952.052
Dec 1993	4504.345	3658.794	5398.417	3232.316	5890.095
Jan 1994	4918.287	4005.015	5883.362	3544.099	6413.884
Feb 1994	4955.395	3995.837	5971.870	3512.727	6531.481

ARRUMAR O ERRO NO ETS_TRANSF

COLOCAR INTERPRETAÇÃO

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_temporais