

Luiz Gustavo Jordão Graciano

Análise de Séries Temporais
Seleção de modelos ARIMA e ETS com foco no
estudo de desempenho preditivo

Brasília, DF

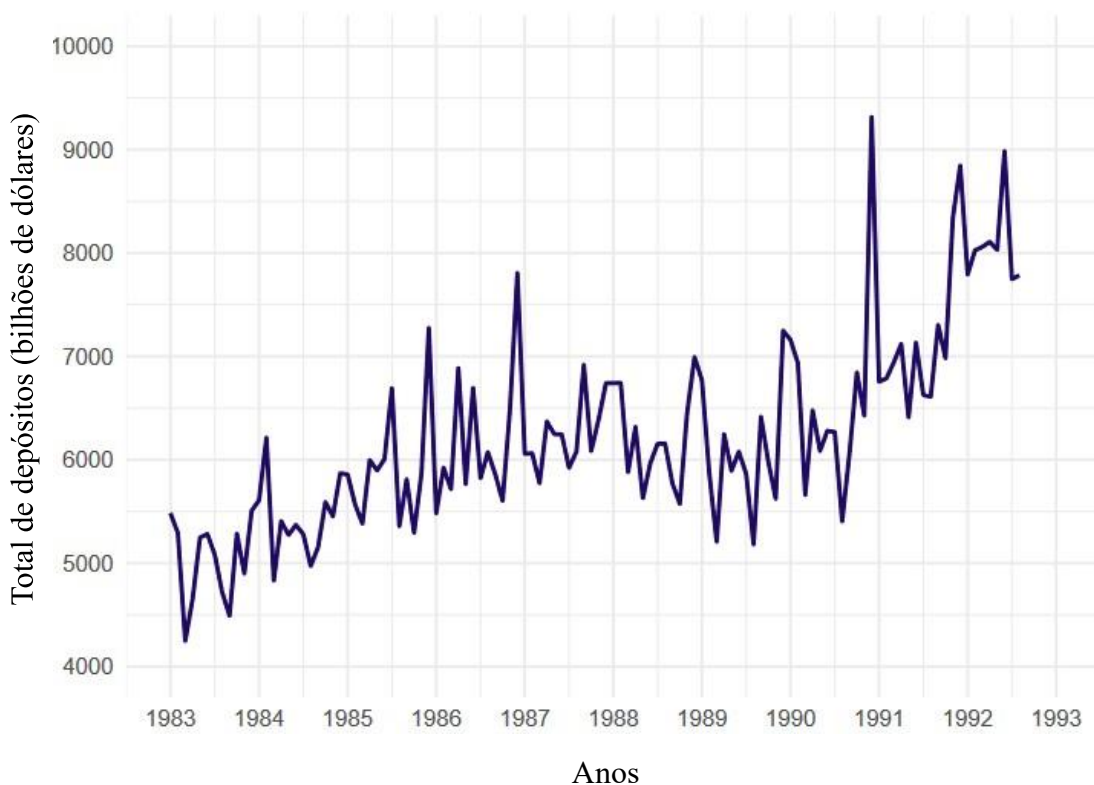
12 de julho de 2023

1. Introdução

O presente trabalho tem como objetivo estudar uma série temporal selecionada do banco de dados da competição de previsão M3, disponibilizado no pacote *Mcomp* do R. Em um primeiro momento, será realizada uma decomposição da série temporal via MSTL e, posteriormente, modelos ARIMA e ETS serão selecionados para realizarmos previsões e um estudo de desempenho preditivo. Em todos os testes estatísticos realizados no estudo, definimos um nível de significância de 0,05.

A série temporal escolhida foi a de número 2539. As 116 observações disponibilizadas se referem ao período entre janeiro de 1983 e agosto de 1992. Os valores da série indicam o total de depósitos à vista, em bilhões de dólares, nos grandes bancos comerciais dos estados americanos e subregiões. Essa categoria de depósito envolve fundos depositados que podem ser sacados a qualquer momento, sem aviso prévio. A série temporal que será estudada pode ser visualizada a seguir:

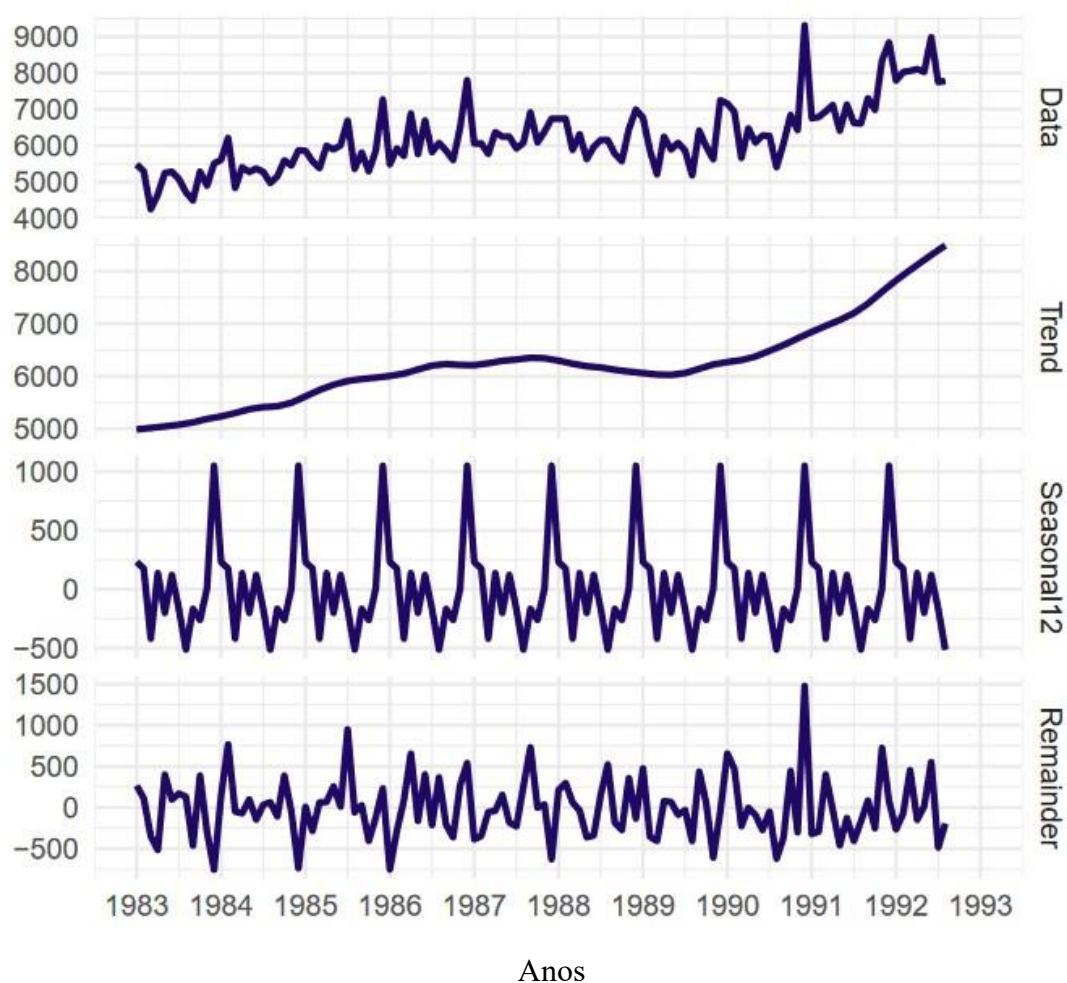
Figura 1 – Série temporal do total de depósitos à vista, em bilhões de dólares, nos grandes bancos comerciais dos estados americanos e subregiões - EUA, 1983-1992



Fonte: International Institute of Forecasters – M3 Competiton

A decomposição MSTL nos permite dividir a estrutura da série temporal em três componentes: tendência, sazonalidade e resíduos. Nesse sentido, obtivemos o seguinte resultado:

Figura 2 – Decomposição via MSTL da série original em componentes de tendência, sazonalidade e resíduos - Estados Unidos, 1983-1992



Fonte: International Institute of Forecasters – M3 Competiton

Conforme esperado, a tendência de crescimento foi confirmada, entretanto, esse comportamento não ocorreu de forma linear. A componente sazonal também foi analisada via MSTL e foi possível captar apenas uma leve tendência geral de pico no mês de janeiro e queda com oscilação nos meses seguintes, todavia, não podemos considerar que o padrão que visualizamos no gráfico ocorre de forma constante. Em síntese, a sazonalidade não está bem definida na série, ou seja, não há um padrão claro.

Por fim, obtemos a componente de resíduos. Podemos observar que os erros apresentam um padrão mais ou menos aleatório, logo, não aparentam estar correlacionados. Nesse sentido, os resíduos incorporaram muito pouca informação de tendência e não há muitos resquícios de sazonalidade, por exemplo. Além disso, os erros estão em torno da média zero. Portanto, a série não aparenta ser estacionária, possui tendência e seu comportamento sazonal não é bem definido.

Observações: Foram testados alguns valores para as janelas de ajuste para sazonalidade e tendência, entretanto, o resultado mais satisfatório e adequado para a decomposição ocorre com o parâmetro *s.window = "periodic"*. Cabe observar que a decomposição também foi realizada a partir do logaritmo da série, uma vez que tentamos obter uma decomposição multiplicativa. Entretanto, o resultado é praticamente o mesmo do que obtivemos acima, embora os resíduos encontrados tenham apresentado comportamento aleatório levemente melhor.

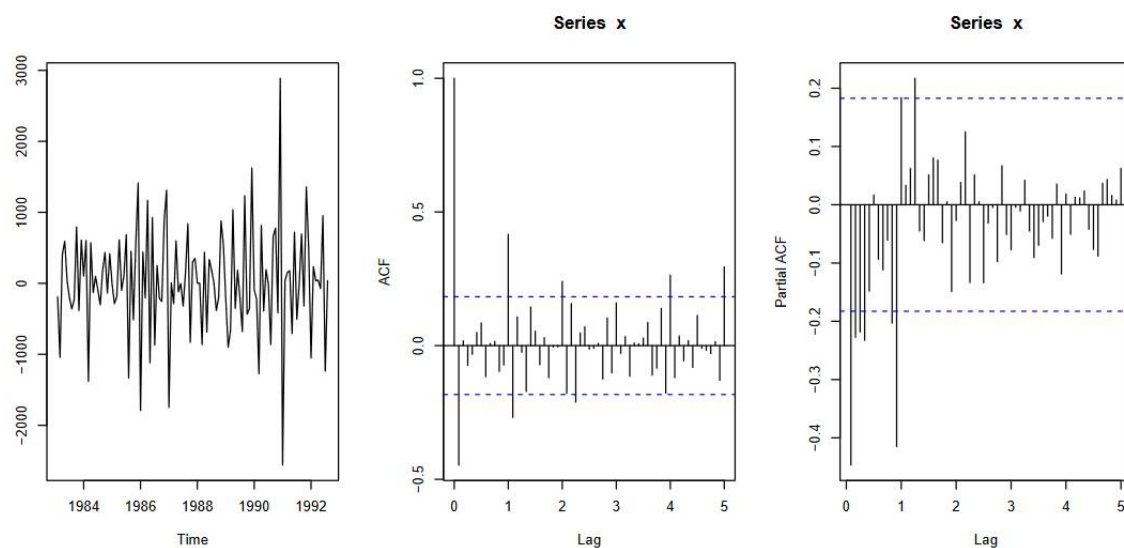
2. Modelos ARIMA

Nessa etapa fizemos a seleção de 2 modelos ARIMA, um sem transformação de Box-Cox e outro com a transformação de Box-Cox. Em ambos os casos o processo de seleção do modelo é semelhante. Podemos representar o algoritmo para a escolha dos modelos da seguinte forma (método de Box-Jenkins): a) conferir a necessidade em remover raízes unitárias; b) identificar candidatos a modelos ARIMA, considerando a interpretação dos gráficos ACF e PACF; c) seleção dos valores de p , q , P e Q , conforme os seguintes critérios: número de parâmetros, AICc e BIC; d) análise de resíduos.

2.1 Modelo ARIMA sem transformação de Box-Cox

Conforme análise acima e fazendo o teste KPSS, obtêm-se um p -valor abaixo de 0,01. De modo que, rejeita-se a hipótese nula de estacionaridade. De fato, podemos utilizar a função **ndiffs** e verificar que é necessário uma diferença para remover as raízes unitárias simples. E em seguida, com a utilização da função **nsdiffs**, notamos que não é necessário remover as raízes sazonais. Embora o modelo não apresente diferença sazonal, ainda precisamos verificar possíveis valores para P e Q . A série diferenciada e os gráficos ACF e PACF podem ser observados abaixo:

Figura 3 – Diferenciação da série original após uma diferença simples, ACF e PACF da série diferenciada - Estados Unidos, 1983-1992



Fonte: International Institute of Forecasters – M3 Competiton

Percebe-se que, de fato, a série atingiu a estacionaridade após a diferenciação. A partir do gráfico de autocorrelações ACF, visualizamos que a série não apresenta correlações que decaem para zero de forma amortizada, entretanto, também não é possível afirmar que existe uma quebra nos lags simples. Nos lags sazonais, há presença de autocorrelação significativa, mas o padrão de queda também não é claro. No gráfico de autocorrelações parciais PACF também não há um padrão muito definido em relação aos lags simples e sazonais. Nessa perspectiva, fizemos uma análise mais detalhada para diferentes valores de p, q, P e Q. O melhor modelo, de acordo com os critérios estabelecidos, foi o **SARIMA(1,1,3)x(1,0,1)₁₂**. A equação encontrada, utilizando os parâmetros estimados e os operadores de retardo, possui a seguinte configuração:

$$(1 - 0,96B^{12})(1 - 0,95B)(1 - B) X_t = (1 - 0,76B^{12})(1 - 1,83B + 0,77B^2 + 0,107B^3) \varepsilon_t$$

em que $\{\varepsilon_t\}$ é um processo i.i.d. com média igual a 0 e desvio-padrão de 478,42. Para a inicialização do modelo precisamos considerar que o termo passado mais distante será o ε_{t-15} , logo, precisamos de $t > 15$.

Tabela 1 – Teste de estacionariedade para as séries original e diferenciada

| Série Temporal | KPSS Level | P-valor | Estacionariedade |
|-------------------|------------|---------|------------------|
| Original | 1,697 | < 0,01 | Rejeitada |
| Diferença simples | 0,039 | > 0,10 | Aceita |

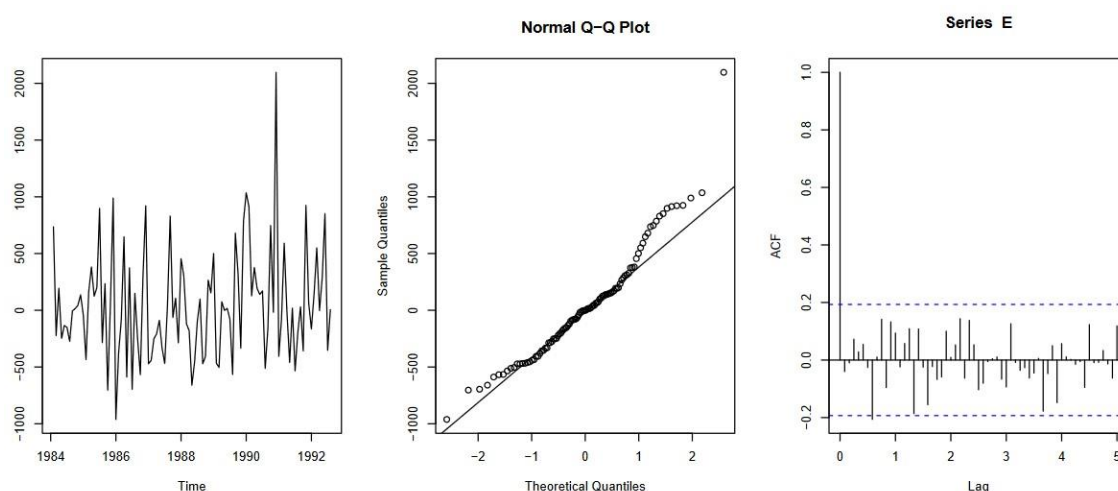
Tabela 2 – Comparação de modelos ARIMA pelo critério AICc e métrica MAE

| Modelos | AIC Corrigido | MAE |
|---|---------------|--------|
| SARIMA(1,1,2)x(0,0,2) ₁₂ | 1773,68 | 411,15 |
| SARIMA(0,1,1)x(1,0,1) ₁₂ | 1771,87 | 491,75 |
| SARIMA(1,1,3)x(1,0,1)₁₂ | 1769,29 | 366,95 |

Para finalizar o processo de escolha do modelo **SARIMA(1,1,3)x(1,0,1)₁₂**, realizamos uma análise de resíduos. O foco nessa etapa está em observar se o modelo está bem ajustado, considerando que as hipóteses levantadas sobre os erros $\{\varepsilon_t\}$ devem ser

atendidas: média em torno do zero, variância constante, autocorrelação nula e normalidade. Além disso, também testamos a estacionariedade.

Figura 4 – Resíduos ajustados, QQ Plot para teste de normalidade e ACF dos resíduos do modelo $SARIMA(1,1,3) \times (1,0,1)_{12}$ - Estados Unidos, 1983-1992



Fonte: International Institute of Forecasters – M3 Competiton

Observando os resultados acima, os erros oscilam em torno de zero. Não se vê nenhum comportamento significativo que aponte que a variância esteja correlacionada com o tempo. A comparação dos quantis da distribuição normal com os pontos distribuídos dos erros mostra uma distância muito grande entre as duas distribuições. No gráfico de autocorrelação, por sua vez, para as defasagens diferentes de zero, o comportamento geral é a autocorrelação contida dentro do intervalo. Não há indícios gráficos de autocorrelação entre os resíduos. Em suma, a normalidade foi rejeitada, entretanto, a independência, a estacionariedade e a inexistência de autocorrelação estão todas satisfeitas na análise dos resíduos. Os resultados dos testes são apresentados abaixo:

Tabela 3 – Resultados da análise de resíduos do modelo $SARIMA(1,1,3) \times (1,0,1)_{12}$

| Teste | Estatística do Teste | P-valor | Situação |
|------------------------------|----------------------|---------|-----------|
| Estacionariedade | KPSS Level = 0,29 | > 0,10 | Aceita |
| Normalidade: Shapiro-Wilk | W = 0,94 | < 0,001 | Rejeitada |

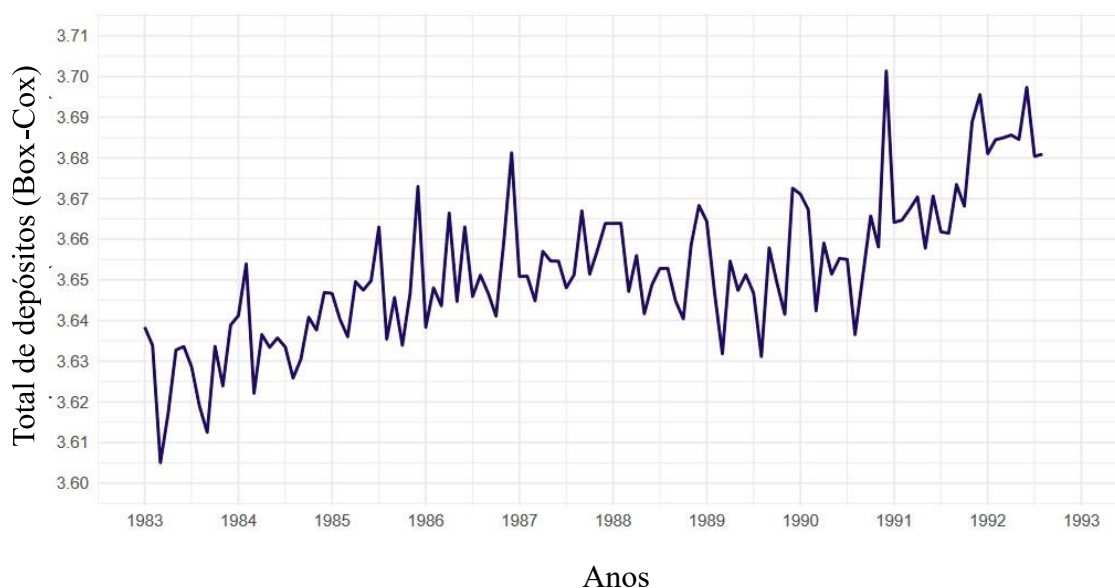
| | | | |
|-----------------------|-------------------|------|--------|
| Independência: Ljung- | X-squared = 14,46 | 0,49 | Aceita |
| Box com lag 15 e 20 | X-squared = 23,50 | 0,26 | Aceita |

2.2 Modelo ARIMA com transformação de Box-Cox

Para a transformação de Box-Cox precisamos definir qual o valor do parâmetro λ . Pelo método automático de seleção no R o valor gerado foi de -0,24. Para uma análise mais sólida, também comparamos os resultados obtidos com outros valores clássicos como $\lambda = 0$, $\lambda = -0,5$, $\lambda = 0,5$, $\lambda = -1$, $\lambda = 1$ e $\lambda = 0,25$. Em termos de seleção do modelo, decomposição da série, AICc e análise de resíduos, o valor final de escolha foi aquele gerado pelo método automático, portanto, temos $\lambda = -0,24$.

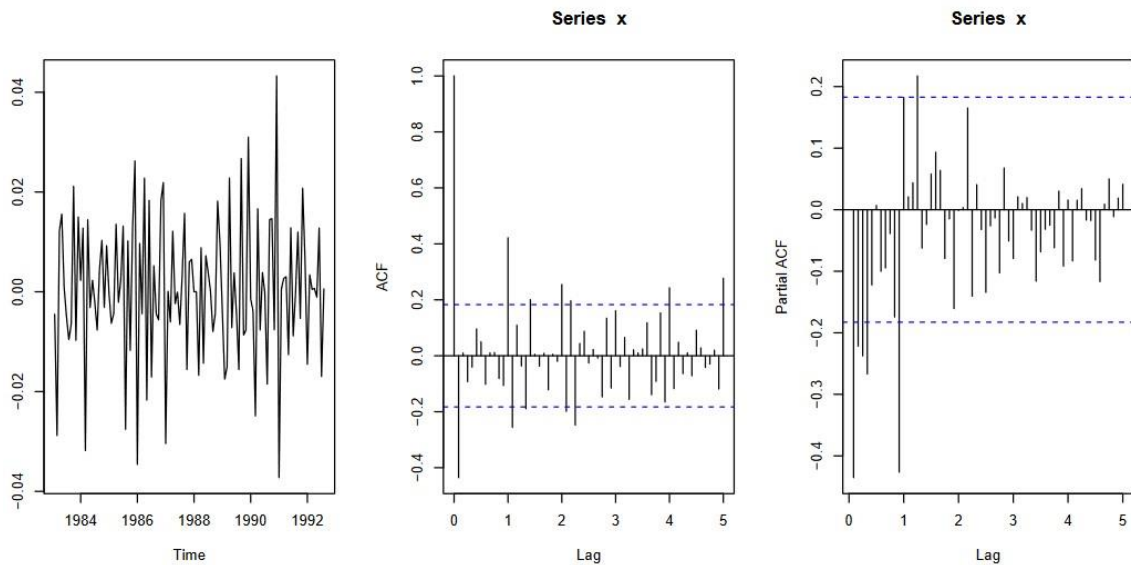
Em termos gerais, os resultados e as interpretações oriundos do modelo com transformação de Box-Cox foram semelhantes ao que foi obtido no item anterior. Portanto, apresentamos à seguir apenas os gráficos e as conclusões gerais, visto que o processo de seleção do modelo e as justificativas são as mesmas. Apesar disso, algumas diferenças foram constatadas e, conseqüentemente, breves comentários serão feitos.

Figura 5 – Série temporal transformada do total de depósitos à vista, em bilhões de dólares, nos grandes bancos comerciais dos estados americanos e subregiões - EUA,



Fonte: International Institute of Forecasters – M3 Competition.

Figura 6 – Diferenciação da série original transformada após uma diferença simples, ACF e PACF da série diferenciada - Estados Unidos, 1983-1992



Fonte: International Institute of Forecasters – M3 Competiton

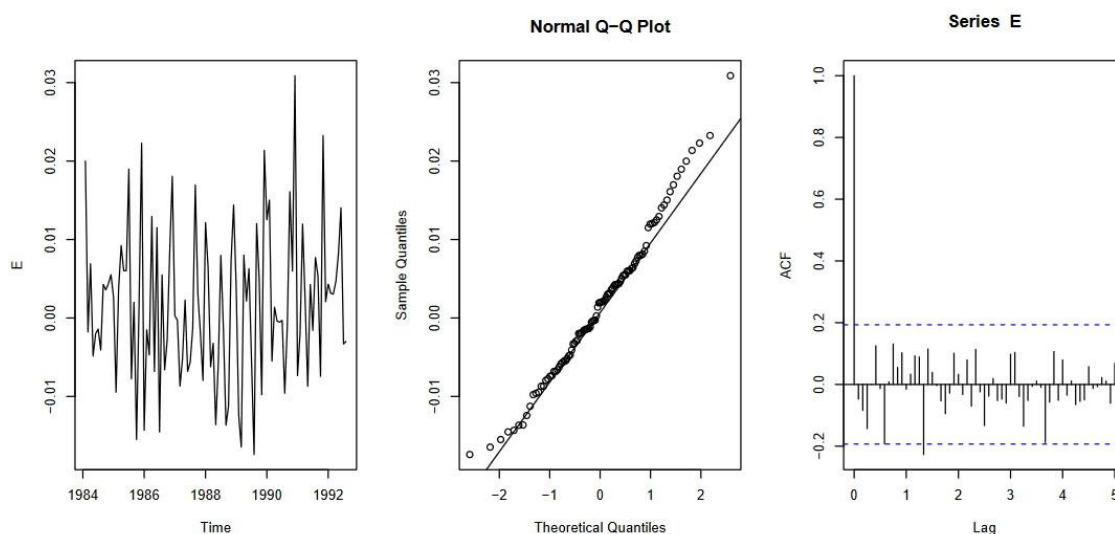
O modelo escolhido será **SARIMA(1,1,2)x(0,0,2)₁₂**, com a seguinte configuração:

$$(1 - 0,93B)(1 - B) X_t = (1 + 0,45B^{12} + 0,36B^{24})(1 - 1,87B + 0,90B^2) \varepsilon_t$$

em que $\{\varepsilon_t\}$ é um processo i.i.d. com distribuição normal e com média igual a 0 e desvio-padrão de 0,00956.

Para finalizar o processo de escolha do modelo **SARIMA(1,1,2)x(0,0,2)₁₂**, realizamos uma análise de resíduos. Em suma, a normalidade, a independência, a estacionariedade e a inexistência de autocorrelação estão todas satisfeitas na análise dos resíduos. Nessa perspectiva, pelo gráfico QQ Plot e teste de normalidade percebemos que a transformação de Box-Cox com $\lambda = -0,24$ nos permitiu obter resíduos que seguem uma distribuição normal, o que não ocorreu no caso anterior. Além disso, o modelo ARIMA selecionado mudou um pouco sua estrutura, embora ainda tenham preservado a necessidade de uma diferença simples e nenhuma diferença sazonal. Os resultados dos testes são apresentados logo abaixo:

Figura 7 – Resíduos ajustados, QQ Plot para teste de normalidade e ACF dos resíduos do modelo $\text{SARIMA}(1,1,2)\times(0,0,2)_{12}$ - Estados Unidos, 1983-1992



Fonte: International Institute of Forecasters – M3 Competiton

Tabela 4 – Teste de estacionariedade para as séries original e diferenciada – Box-Cox

| Série Temporal | KPSS Level | P-valor | Estacionariedade |
|-------------------|------------|---------|------------------|
| Original | 1,750 | < 0,01 | Rejeitada |
| Diferença simples | 0,038 | > 0,10 | Aceita |

Tabela 5 – Comparação de modelos pelo critério AICc e métrica MAE – Box-Cox

| Modelos | AIC Corrigido | MAE |
|--|---------------|--------|
| $\text{SARIMA}(1,1,3)\times(0,0,1)_{12}$ | -717.27 | 470,29 |
| $\text{SARIMA}(0,1,1)\times(0,0,2)_{12}$ | -725.48 | 519,50 |
| $\text{SARIMA}(1,1,2)\times(0,0,2)_{12}$ | -727.47 | 442,32 |

Tabela 6 – Resultados da análise de resíduos do modelo **SARIMA(1,1,2)x(0,0,2)₁₂**

| Teste | Estatística do Teste | P-valor | Situação |
|---|--|--------------|------------------|
| Estacionariedade | KPSS Level = 0,15 | > 0,10 | Aceita |
| Normalidade: Shapiro-Wilk | W = 0,98 | 0,37 | Aceita |
| Independência: Ljung-Box com lags 15 e 20 | X-squared = 10,37 X-squared = 18,62 | 0,80 0,55 | Aceita Aceita |

3. Modelos ETS

Nessa etapa fizemos a seleção de 2 modelos ETS, um sem transformação de BoxCox e outro com a transformação de Box-Cox. Em ambos os casos o processo de seleção do modelo é semelhante e envolve, em grande medida, o estudo que fizemos sobre a estrutura da série temporal.

3.1 Modelo ETS sem transformação de Box-Cox

Para escolha adequada de um modelo ETS, é preciso escolher se o comportamento dos erros, tendência e sazonalidade existem na série, caso existam, então segue-se escolhendo um modelo que os descrevem de forma aditiva ou multiplicativa.

Utilizando a decomposição da série, identifica-se que há tendência na série e que o comportamento da tendência aparenta ser da forma aditiva, porém não é constante, tornando necessária a estimação de um parâmetro de damped. Da decomposição, também é possível observar que há uma característica sazonal na série (embora não muito bem definida), contudo, o gráfico da série não deixa claro se a sazonalidade é aditiva ou multiplicativa. Um ponto que sustenta essa dúvida é que a decomposição multiplicativa também captou a sazonalidade e apresentou erros similares a um ruído branco, portanto, seguiu-se comparando os Critérios corrigidos de Akaike como indicado pela tabela abaixo:

Tabela 7: Critérios de seleção de modelos para modelos candidatos ETS.

| ETS | AICc |
|------|----------|
| AAdA | 1998.441 |
| MAdA | 1993.861 |
| MAdM | 1989.925 |

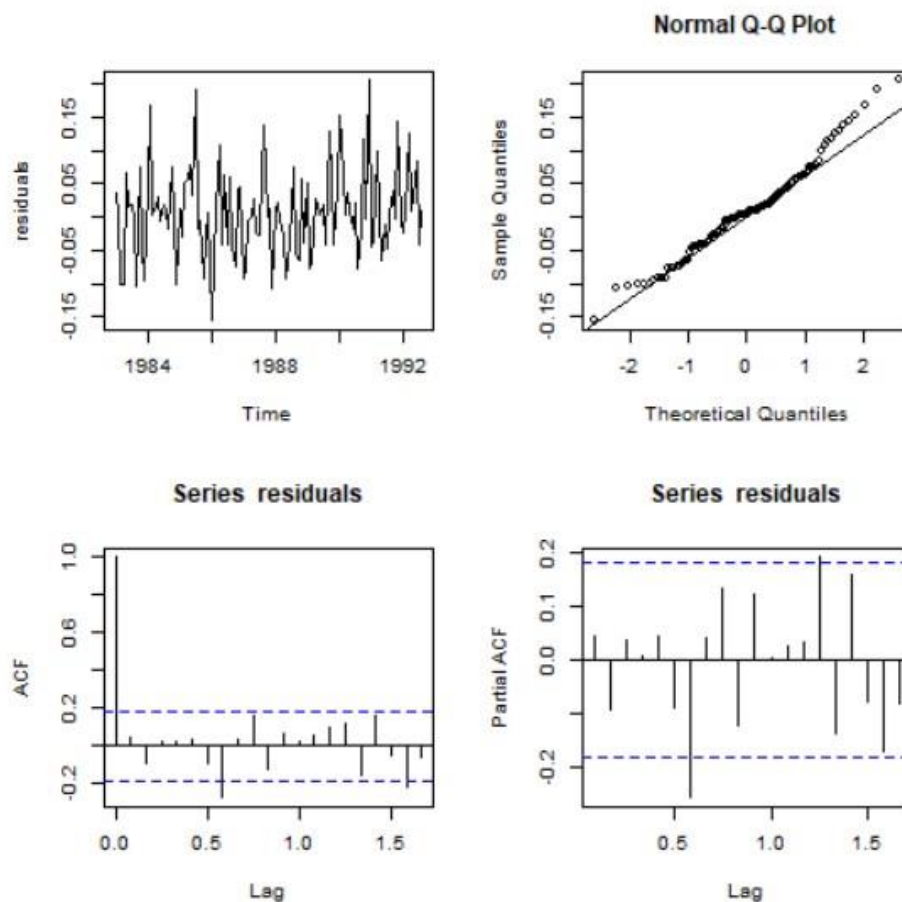
Logo, o modelo escolhido é aquele que apresenta o menor valor de AICc, isto é, um modelo com a seguinte configuração:

$$ETS(M, Ad, M)$$

Procedendo à análise de resíduos do modelo, tem-se que no gráfico a seguir há ocorrência de resíduos que se parecem com um ruído branco, isto é, parecem se comportar de forma independente, com média nula e variância constante. O qqplot indica que os erros não se ajustam bem à distribuição normal devido à sua variação com respeito à reta de probabilidade normal.

Os gráficos ACF e PACF reforçam que não sobrou correlação significativa entre os erros, os únicos saltos além da faixa que chamam mais atenção nos gráficos ACF e PACF não estão no lag sazonal e estão por bem pouco fora da faixa, levando a crer que o ajuste do modelo está adequado.

Figura 8 – Gráficos de linha, probabilidade normal, ACF e PACF para os resíduos do modelo $ETS(M, Ad, M)$ sem transformação de Box-Cox - Estados Unidos, 1983-1992



Para apoiar a análise gráfica, a tabela 8 apresenta testes de hipóteses, suas respectivas estatísticas do teste e p-valor, que nos levam a conclusão de que os resíduos se comportam de forma estacionária, são independentes entre si e não seguem distribuição normal. Vale notar que para o teste de Ljung-Box foi utilizado lag igual a 15.

Tabela 8: Teste de hipóteses para pressupostos do modelo $ETS(M, Ad, M)$ sem a transformação de Box-Cox

| Estatística | Teste | Valor da Estatística | P-valor |
|-------------|--------------|----------------------|---------|
| KPSS Trend | KPSS | 0.070 | > 0.1 |
| X-squared | Ljung-Box | 21.682 | 0,116 |
| W | Shapiro-Wilk | 0.977 | 0,042 |

3.2 Modelo ETS com transformação de Box-Cox

Utilizando o mesmo parâmetro para a transformação de Box-Cox, escolhido na seção de modelos ARIMA ($\lambda = -0.24$), tem-se que os menores valores da série transformada se estabelecem muito próximos de zero, por conseguinte não é adequado trabalhar com componente de erro multiplicativa, por conta dos chamados “modelos instáveis”. Lembrando que a série apresenta tendência e devemos considerar a sazonalidade (a série modelada por ETS com sazonalidade apresenta resultados mais satisfatórios), então, o único modelo possível para ser ajustado pela função *ets()* da biblioteca *forecast* em R é com todas as componentes aditivas. Para este ajuste, o parâmetro de damped ajustado foi próximo de 1, concluindo a escolha do modelo HoltWinters Aditivo + damped:

$$ETS(A, Ad, A)$$

De forma análoga ao modelo ETS sem a transformação de Box-Cox, tem-se na figura e na tabela abaixo a indicação de que o modelo com a transformação de Box-Cox atende todos os pressupostos, uma vez que seu gráfico de linha para os resíduos se parece com um ruído branco. Pelo teste de Ljung-Box com lag igual a 15 conclui-se que há independência dos resíduos, ponto que é sustentado pelos gráficos ACF e PACF. Pelo teste de Shapiro-Wilk e gráfico de probabilidade normal conclui-se que os resíduos seguem distribuição normal e pelo teste KPSS os resíduos descrevem um processo estacionário.

Figura 9 – Gráficos de linha, probabilidade normal, ACF e PACF para os resíduos do modelo $ETS(A, Ad, A)$ com transformação de Box-Cox - Estados Unidos, 1983-1992

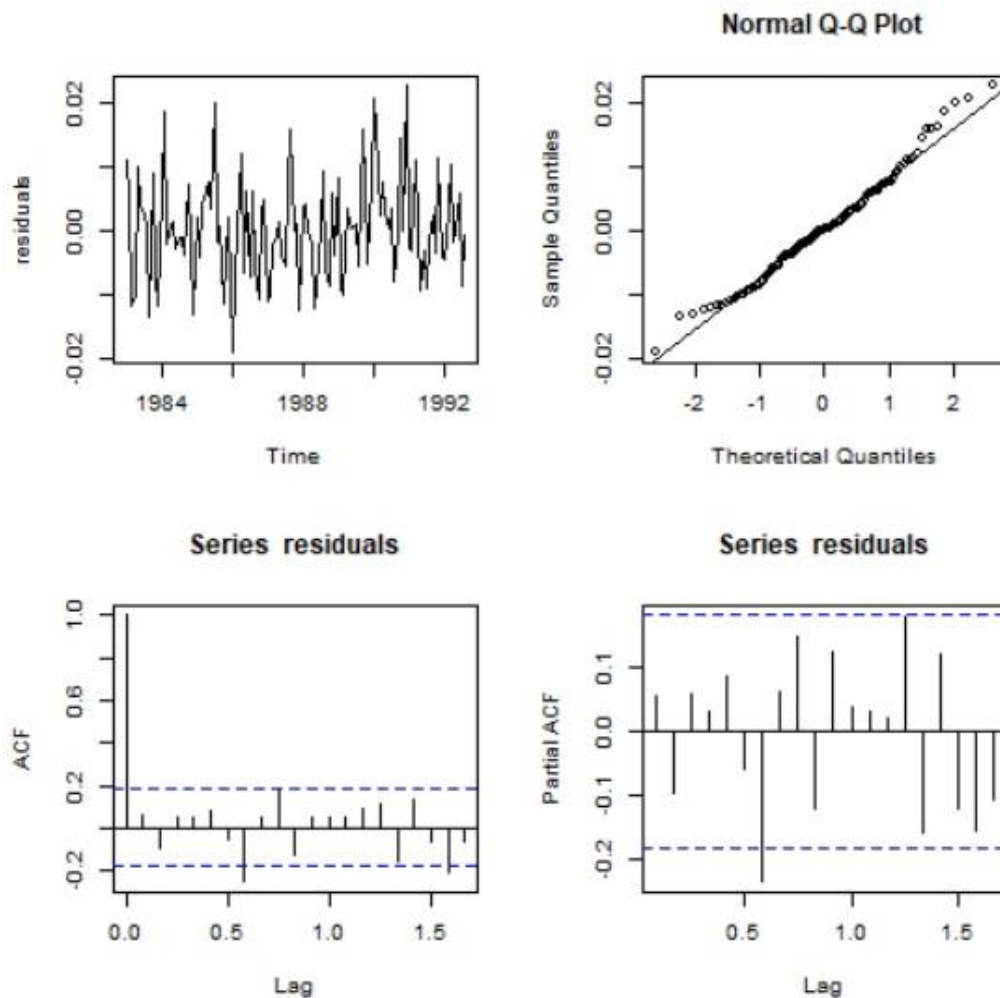


Tabela 9: Teste de hipóteses para pressupostos do modelo $ETS(A, Ad, A)$ com a transformação de Box-Cox

| Estatística | Teste | Valor da Estatística | P-valor |
|-------------|--------------|----------------------|---------|
| KPSS Trend | KPSS | 0,080 | > 0.1 |
| X-squared | Ljung-Box | 21,451 | 0,123 |
| W | Shapiro-Wilk | 0,984 | 0,192 |

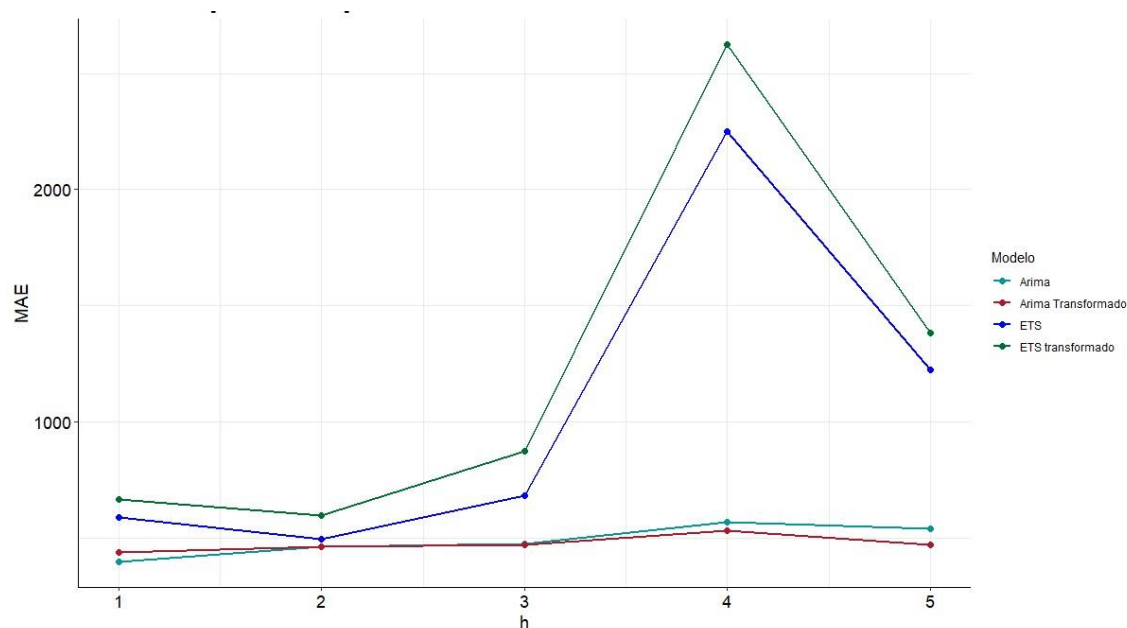
4. Estudo de desempenho preditivo

Para a avaliação do desempenho preditivo por janela deslizante, considera-se o estudo de janela em $n-14$ e os erros para os horizontes até 5. Para cada modelo previamente estabelecido pelos critérios do trabalho, foram construídas as funções de previsões com o intuito de avaliar os erros absolutos médio (MAE) para cada horizonte de previsão. Segue os resultados dos erros absolutos médios para cada modelo:

Tabela 10: MAE de todos os modelos selecionados e respectivos horizontes de previsão.

| Horizonte | Arima | ETS | Arima Transformado | ETS transformado |
|-----------|----------|-----------|--------------------|------------------|
| h=1 | 398.8380 | 590.3297 | 439.6815 | 666.7057 |
| h=2 | 462.1857 | 495.6874 | 462.4981 | 597.3083 |
| h=3 | 474.2587 | 683.1605 | 471.0477 | 872.4148 |
| h=4 | 570.1111 | 2251.6798 | 532.1995 | 2622.6185 |
| h=5 | 539.5040 | 1222.7496 | 468.7280 | 1382.6881 |

Figura 10 – Erro absoluto médio de previsão para os 4 modelos selecionados



Analisando a tabela acima temos que para os horizontes 1 e 2 o modelo **SARIMA(1,1,3)x(1,0,1)₁₂** é considerado o melhor, a partir do critério de que o melhor modelo é aquele que apresenta o menor erro, porém o modelo **SARIMA(1,1,2)x(0,0,2)₁₂** também aparentou ser eficiente e com melhores indicadores para os horizontes 3, 4 e 5. Em suma, nos dois modelos obtivemos resultados próximos.

O modelo **ETS(M, Ad, M)** apresenta resultados menos precisos do que os modelos ARIMA selecionados. Nos horizontes 1, 2 e 3 os erros não são tão significativos, entretanto, para horizontes maiores o erro absoluto médio passa a ser bem mais amplo. Por sua vez, o modelo **ETS(A, Ad, A)** para a série transformada por Box-Cox apresentou o pior desempenho nos 5 horizontes sob estudo.

5. Resultados

Figura 11 – Previsão pontual e previsão intervalar (95%) do modelo $\text{SARIMA}(1,1,3)\times(1,0,1)_{12}$. para 18 passos a frente – sem transformação Box-Cox

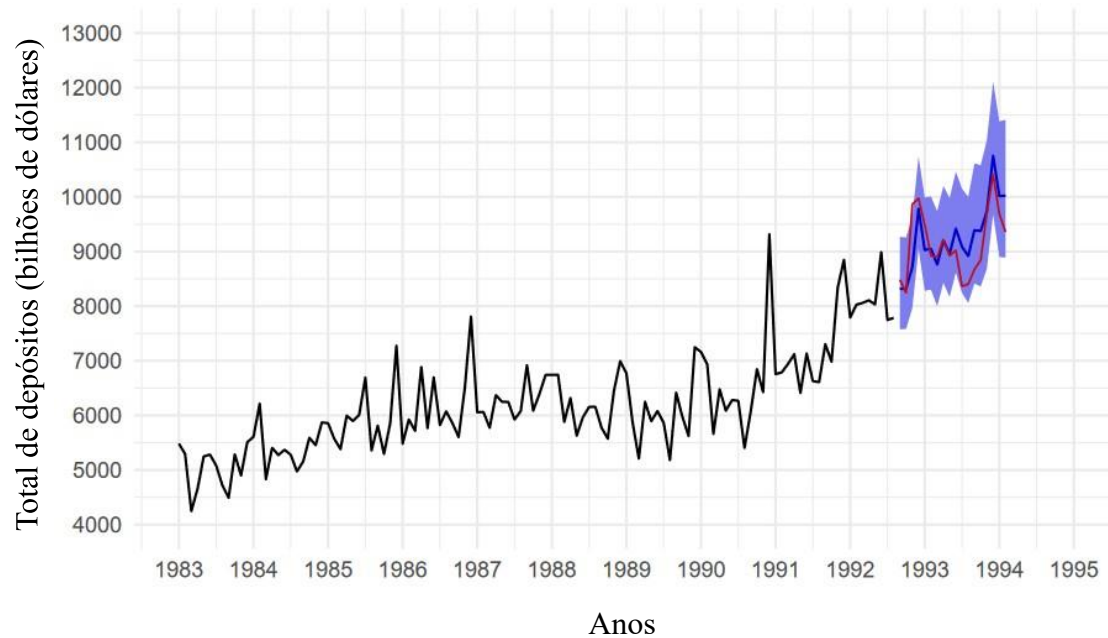
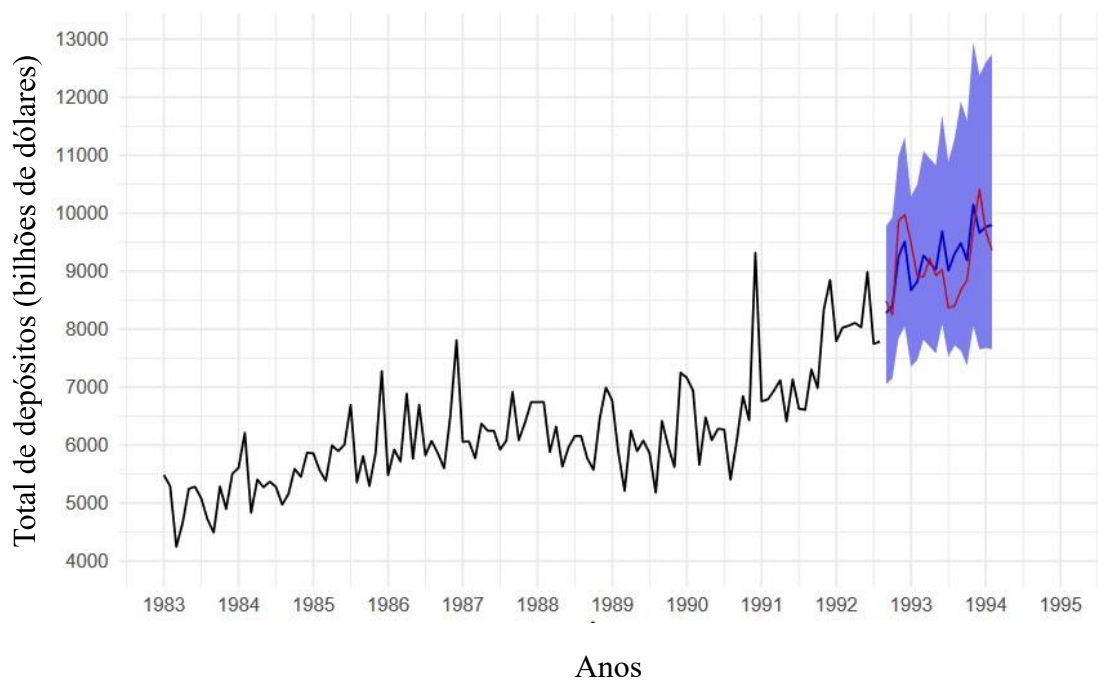


Figura 12 – Previsão pontual e da previsão intervalar (95%) do modelo $\text{SARIMA}(1,1,2)\times(0,0,2)_{12}$. para 18 passos a frente – com transformação Box-Cox



Uma vez que o modelo ARIMA escolhido sem a transformação de Box-Cox falhou para o teste de normalidade, foi feito o uso da técnica de Bootstrap calculando quantis empíricos da amostra de resíduos obtida para compor as previsões intervalares, que podem ser observadas no gráfico da figura 11. O resultado encontrado foi satisfatório, visto que o intervalo obtido possui menor amplitude com uma precisão excelente. A curva vermelha com os valores reais da série se assemelha bastante com a curva azul para as previsões pontuais, além disso, as previsões e os valores verdadeiros caem dentro do intervalo de confiança de 95%.

Para o gráfico da figura 12 tem-se o modelo ARIMA com transformação de BoxCox, que não rejeitou a hipótese de normalidade para o teste de Shapiro-Wilk, portanto foram calculadas a partir dos quantis da distribuição normal padrão. O resultado obtido também foi satisfatório, entretanto, o intervalo de confiança gerado possui maior amplitude. A curva vermelha com os valores reais da série está próxima da curva azul para as previsões pontuais, embora sejam diferentes. As previsões e os valores verdadeiros caem dentro do intervalo de confiança de 95%.

Tabela 11: Previsão pontual e intervalar (95%) para os modelos ARIMA

| Data | | SARIMA(1,1,3)x(1,0,1) ₁₂ | | | SARIMA(1,1,2)x(0,0,2) ₁₂ | | |
|------|------|-------------------------------------|-----------------|-----------------|-------------------------------------|-----------------|-----------------|
| Mês | Ano | Previsão | Limite Inferior | Limite Superior | Previsão | Limite Inferior | Limite Superior |
| Sep | 1992 | 8.315,86 | 7.576,33 | 9.268,91 | 8.275,28 | 7.049,72 | 9.776,43 |
| Oct | 1992 | 8.317,37 | 7.582,14 | 9.244,19 | 8.397,67 | 7.148,45 | 9.929,31 |
| Nov | 1992 | 8.714,92 | 7.975,87 | 9.645,97 | 9.255,30 | 7.845,49 | 10.993,25 |
| Dec | 1992 | 9.781,12 | 9.041,79 | 10.718,82 | 9.507,67 | 8.044,68 | 11.315,47 |
| Jan | 1993 | 9.027,17 | 8.281,10 | 10.004,14 | 8.668,48 | 7.352,86 | 10.288,96 |
| Feb | 1993 | 9.050,11 | 8.290,24 | 10.033,49 | 8.818,86 | 7.464,31 | 10.491,95 |
| Mar | 1993 | 8.762,32 | 7.994,08 | 9.751,11 | 9.269,10 | 7.815,18 | 11.073,93 |
| Apr | 1993 | 9.205,14 | 8.410,65 | 10.194,42 | 9.142,94 | 7.695,50 | 10.943,74 |
| May | 1993 | 8.954,93 | 8.162,89 | 9.997,89 | 9.025,32 | 7.580,38 | 10.827,97 |
| Jun | 1993 | 9.417,46 | 8.611,61 | 10.490,64 | 9.686,48 | 8.087,16 | 11.697,30 |
| Jul | 1993 | 9.083,77 | 8.240,61 | 10.164,73 | 9.008,85 | 7.519,16 | 10.882,56 |
| Aug | 1993 | 8.919,09 | 8.031,15 | 10.047,29 | 9.297,31 | 7.720,97 | 11.293,03 |
| Sep | 1993 | 9.389,30 | 8.427,44 | 10.568,92 | 9.483,47 | 7.627,53 | 11.933,58 |
| Oct | 1993 | 9.374,89 | 8.389,07 | 10.596,80 | 9.187,67 | 7.369,01 | 11.597,36 |
| Nov | 1993 | 9.741,97 | 8.717,05 | 11.068,07 | 10.146,06 | 8.054,63 | 12.954,87 |
| Dec | 1993 | 10.753,22 | 9.702,78 | 12.079,80 | 9.660,75 | 7.646,89 | 12.375,84 |
| Jan | 1994 | 10.013,88 | 8.897,43 | 11.425,16 | 9.762,45 | 7.676,35 | 12.599,69 |

| | | | | | | |
|----------|-----------|----------|-----------|----------|----------|-----------|
| Feb 1994 | 10.022,78 | 8.877,59 | 11.407,02 | 9.793,69 | 7.649,71 | 12.735,49 |
|----------|-----------|----------|-----------|----------|----------|-----------|

Figura 13 – Previsão pontual e previsão intervalar (95%) do modelo **ETS(M, Ad, M)** para 18 passos a frente – sem transformação Box-Cox

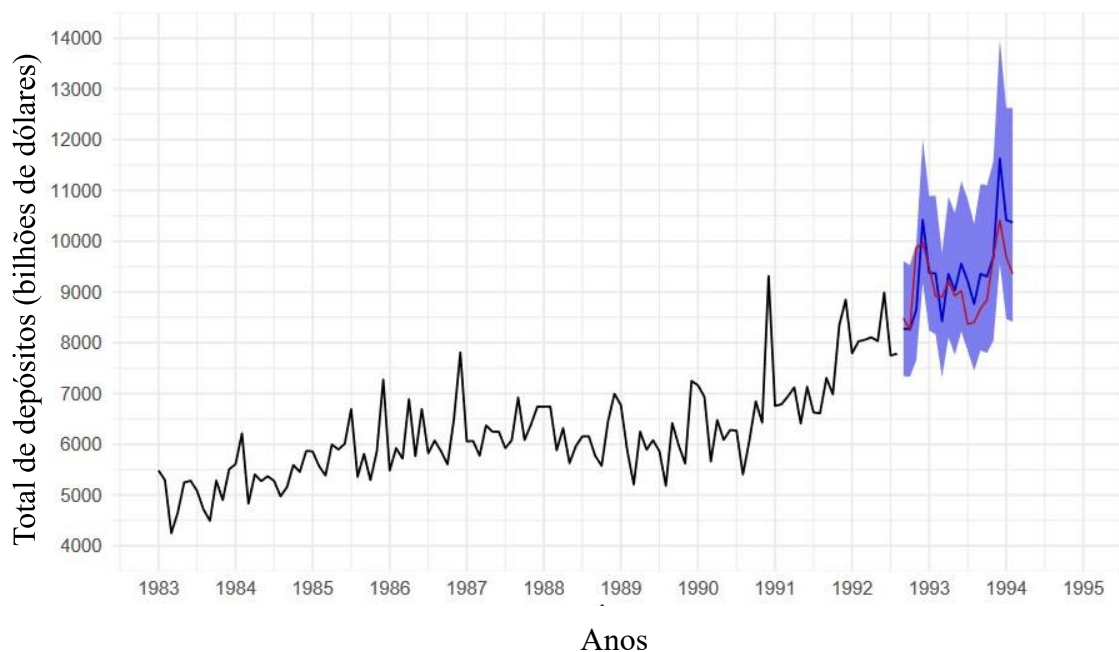
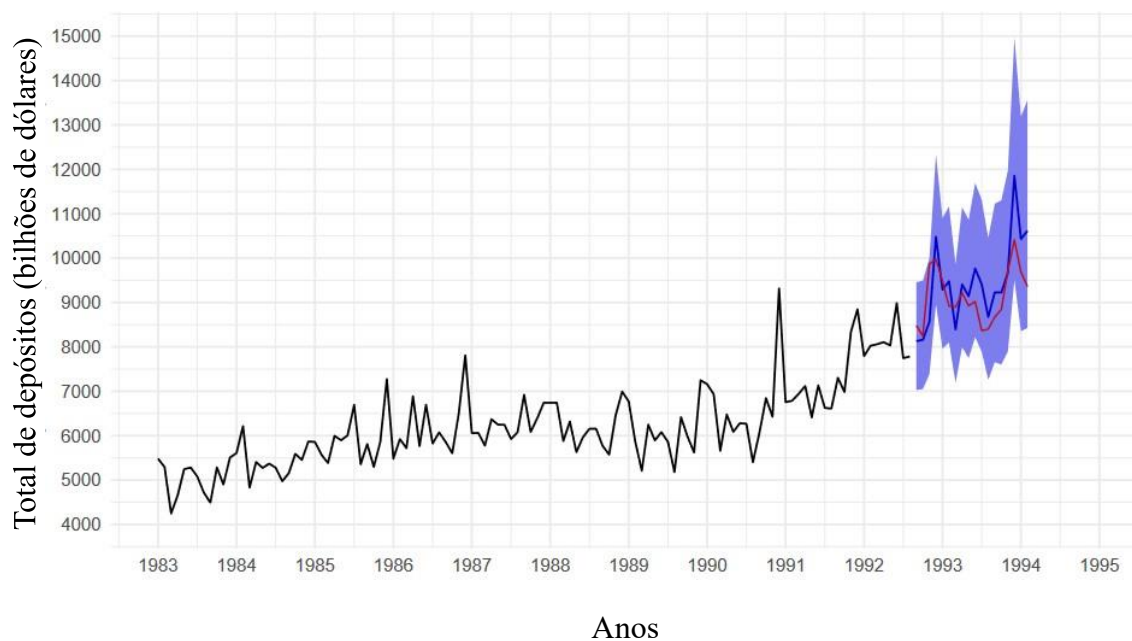


Figura 14 – Previsão pontual e previsão intervalar (95%) do modelo **ETS(A, Ad, A)** para 18 passos a frente – com transformação Box-Cox



Uma vez que o modelo ETS escolhido sem a transformação de Box-Cox falhou para o teste de normalidade, portanto, foi feito o uso da técnica de Bootstrap calculando quantis empíricos da amostra de resíduos obtida para compor as previsões intervalares, que podem ser observadas no gráfico da figura 13. Para o gráfico da figura 14 tem-se o modelo ETS com transformação de Box-Cox, que não rejeitou a hipótese de normalidade para o teste de Shapiro-Wilk, portanto foram calculadas a partir dos quantis da distribuição normal padrão. Ambos os gráficos apresentam uma linha em vermelho que representa o caminho verdadeiro que a série percorreu.

Um ponto interessante é que embora ambas as previsões intervalares para os dois modelos contenham os resultados reais (linha em vermelho), tem-se que o modelo ETS sem a transformação e Box-Cox aparenta ter se saído muito melhor em termos de replicar o comportamento da série e fornecer intervalos mais fechados, mas que ainda sim capturem o verdadeiro valor observado para a série. Em suma, os resultados também foram satisfatórios com os modelos ETS, entretanto, o modelo ARIMA sem transformação de Box-Cox gerou previsões com maior precisão.

Tabela 12: Previsão pontual e intervalar (95%) para os modelos ETS

| Data | | ETS(M, Ad, M) | | | ETS(A, Ad, A) | | |
|------|------|---------------|-----------------|-----------------|---------------|-----------------|-----------------|
| Mês | Ano | Previsão | Limite Inferior | Limite Superior | Previsão | Limite Inferior | Limite Superior |
| Sep | 1992 | 8.274,93 | 7.341,65 | 9.607,64 | 8.131,12 | 7.027,86 | 9.457,48 |
| Oct | 1992 | 8.265,45 | 7.328,98 | 9.527,41 | 8.156,76 | 7.046,32 | 9.492,65 |
| Nov | 1992 | 8.647,94 | 7.649,29 | 9.958,45 | 8.567,05 | 7.382,64 | 9.996,42 |
| Dec | 1992 | 10.419,88 | 9.189,64 | 12.012,48 | 10.479,79 | 8.955,07 | 12.340,00 |
| Jan | 1993 | 9.372,44 | 8.236,01 | 10.888,67 | 9.286,55 | 7.957,55 | 10.902,16 |
| Feb | 1993 | 9.364,94 | 8.167,11 | 10.900,06 | 9.479,06 | 8.098,92 | 11.163,14 |
| Mar | 1993 | 8.421,90 | 7.316,24 | 9.770,38 | 8.394,89 | 7.185,54 | 9.867,13 |
| Apr | 1993 | 9.352,28 | 8.103,98 | 10.873,66 | 9.407,72 | 7.992,20 | 11.147,77 |
| May | 1993 | 9.028,10 | 7.760,21 | 10.558,73 | 9.140,94 | 7.744,33 | 10.863,83 |
| Jun | 1993 | 9.555,19 | 8.221,04 | 11.186,41 | 9.765,94 | 8.215,47 | 11.696,28 |
| Jul | 1993 | 9.210,75 | 7.840,93 | 10.822,37 | 9.410,85 | 7.889,04 | 11.314,14 |
| Aug | 1993 | 8.763,01 | 7.444,81 | 10.339,96 | 8.674,49 | 7.256,51 | 10.452,75 |
| Sep | 1993 | 9.356,63 | 7.850,64 | 11.126,81 | 9.228,55 | 7.653,48 | 11.226,18 |
| Oct | 1993 | 9.303,00 | 7.796,50 | 11.102,05 | 9.224,25 | 7.600,40 | 11.301,25 |
| Nov | 1993 | 9.690,92 | 8.038,31 | 11.574,94 | 9.667,28 | 7.892,68 | 11.964,46 |
| Dec | 1993 | 11.627,83 | 9.565,46 | 13.974,43 | 11.853,66 | 9.507,72 | 14.961,82 |

| | | | | | | | |
|-----|------|-----------|----------|-----------|-----------|----------|-----------|
| Jan | 1994 | 10.417,29 | 8.469,10 | 12.625,06 | 10.429,93 | 8.353,56 | 13.186,21 |
| Feb | 1994 | 10.369,35 | 8.414,68 | 12.626,84 | 10.616,03 | 8.424,21 | 13.561,27 |

MAE: Erro médio absoluto

As métricas MAE para cada um dos modelos implementados e os *benchmarks* da biblioteca *forecast* estão apresentados na tabela abaixo, de onde observa-se que os modelos com melhores desempenhos preditivos segundo o MAE foram a escolha manual do **SARIMA(1, 1, 3)x(1, 0, 1){12}**, o *auto.arima* e a escolha manual do **SARIMA(1, 1, 2)x(0, 0, 2){12}** com a transformação de Box-Cox, nessa ordem, do melhor para o terceiro mais eficiente.

Outro ponto que vale destaque é que a escolha de forma manual do modelo **ETS(M, Ad, M)** concidiu com o *benchmark* automático *ets* da biblioteca *forecast* e que de fato a escolha do modelo sem a transformação de Box-Cox se saiu melhor que a versão com a transformação de Box-Cox, suspeita que já tinha sido instaurada durante a análise gráfica e tabular das previsões pontuais e intervalares.

Tabela 13: Erro médio absoluto para as previsões de 18 passos a frente.

| Implementação | MAE |
|-------------------------|----------|
| ARIMA (manual) | 366,96 |
| ARIMA (manual)(Box-Cox) | 442,32 |
| ETS (manual) | 501,80 |
| ETS (manual)(Box-Cox) | 578,03 |
| <i>auto.arima</i> | 411,15 |
| <i>ses</i> | 1.095,08 |
| <i>holt</i> | 904,89 |
| <i>ets</i> | 503,22 |
| <i>stlf</i> | 556,73 |
| <i>bats</i> | 968,30 |
| <i>tbats</i> | 790,82 |

6. Conclusão

A série selecionada não aparenta ser estacionária, possui tendência e seu comportamento sazonal não é constante. Os modelos ARIMA selecionados sem e com transformação de BoxCox foram o **SARIMA(1,1,3)_x(1,0,1)₁₂** e o **SARIMA(1,1,2)_x(0,0,2)₁₂**, respectivamente. Em relação aos modelos ETS selecionados sem e com transformação de Box-Cox, temos os modelos **ETS(M, Ad, M)** e **ETS(A, Ad, A)**, respectivamente.

Em suma, os modelos ARIMA selecionados apresentaram os menores erros absolutos médios para previsão em comparação com os modelos ETS, indiciando que nesse caso um modelo ARIMA seja o favorito para a escolha. Essa afirmação também é validada pelo estudo de janela deslizando e pelos gráficos e tabelas com as previsões pontuais, intervalares e os valores reais da série. Cabe observar que em comparação com os *benchmarks* da biblioteca *forecast* os modelos ARIMA também indicaram melhores resultados. Portanto, pelo fato de que o modelo ARIMA sem transformação obteve o melhor comportamento em relação ao desempenho preditivo e a melhor acurácia entre todos os modelos apresentados, é justo selecionar tal modelo como o mais adequado.

ANEXO

Código em R

12-07-2023

```
# PACOTES UTILIZADOS E SÉRIE TEMPORAL ESCOLHIDA
```

```
library(Mcomp) library(forecast) library(tidyverse)
```

```
data(M3) id=2539 # id da série  
escolhida
```

```
horizonte <- M3[[id]]$h serie_temp <-  
M3[[id]]$x out_sample <- M3[[id]]$xx
```

```
M3[[id]]$description # descrição da série
```

```
M3[[id]]$type  
M3[[id]]$period  
M3[[id]]$n
```

```
# Gráfico da Série serie_temp %>% autoplot() +  
geom_line(color = "#1f0866", linewidth = 0.8) + xlab("Anos") +  
ylab("Depósitos Totais (bilhões de dólares)") +  
  scale_x_continuous(breaks = seq(from = 1983,to = 1993,by = 1), limits = c(1983,1993))  
  +  
  scale_y_continuous(breaks = seq(from = 4000,to = 10000,by = 1000), limits = c(4000,10000))  
  +  
  theme(axis.text.x=element_text(angle=0, hjust=1)) +  
theme_minimal() serie_temp %>% plot() # opção mais simples
```

```
# DECOMPOSIÇÃO DA SÉRIE TEMPORAL VIA MSTL
```

```
mstl(serie_temp, s.window = "periodic") %>% autoplot() + geom_line(color = c("#1f0866"),  
  linewidth = 1) +
```



```

scale_x_continuous(breaks = seq(from = 1983,to = 1993,by = 1), limits = c(1983,1993))
+
theme_minimal() mstl(serie_temp, s.window = "periodic", lambda = 0) %>%

plot()

# MODELO ARIMA - SEM TRANSFORMAÇÃO DE BOX-COX

# verificando estacionariedade tseries::kpss.test(serie_temp)

# Operador de diferença (tendencia e sazonalidade) ndiffs(serie_temp) #
1 diferença      d = 1 diff(serie_temp) %>% nsdiffs() # 0 diferença D = 0

# SARIMA(p,1,q)x(P,0,Q) x <- diff(serie_temp) # Série estacionária
par(mfrow=c(1,3)) plot(x);acf(x,lag.max = 12*5);pacf(x, lag.max =
12*5) tseries::kpss.test(x) par(mfrow=c(1,1))

# VERIFICANDO MODELOS ADEQUADOS VIA CRITÉRIO AIC CORRIGIDO
melhor_AICc = Inf for(P in 0:2){
  for(Q in 0:2){
    for(p in 0:3){
      for(q in 0:3){ fit = Arima(serie_temp, order = c(p,1,q), seasonal = c(P,0,Q))
        if(fit$aicc < melhor_AICc){ melhor_AICc = fit$aicc cat("p =",p,"q =",q,"P =",P,"Q
        =",Q,
          ",AICc =",fit$aicc, "\n") }
      }
    }
  }
}

# ANÁLISE DE RESÍDUOS DO MODELO 1 ARIMA

# Modelo SARIMA(1,1,3)x(1,0,1) fit = Arima(serie_temp, order=c(1,1,3),
seasonal=c(1,0,1)) fit

# Resíduos
par(mfrow=c(1,2)) E_1 <-
fit$residuals

```



```

plot(E_1)
E <- fit$residuals %>% window(start=c(1984,2)) plot(E)

# gráficos
par(mfrow=c(1,3)) plot(E)
qqnorm(E); qqline(E) acf(E,
lag.max = 12*5) hist(E)

# verificando estacionariedade tseries::kpss.test(E)

# verificando independência
Box.test(E,lag = 15, type = "Ljung-Box")
Box.test(E,lag = 20, type = "Ljung-Box")

# Normalidade dos resíduos shapiro.test(E)

# Horizonte de previsão par(mfrow=c(1,1)) prev = forecast(fit, horizonte,
level = 95, bootstrap = T) prev

library(zoo) d1 = out_sample %>%
as.numeric() d2 =
as.yearmon(time(out_sample))

prev %>% autoplot() + geom_line(linewidth = 0.8) + xlab("Anos") +
  ylab("Depósitos Totais (bilhões de dólares)") + scale_x_continuous(breaks =
seq(from = 1983,to = 1995,by = 1), limits = c(1983,1995)) +
  scale_y_continuous(breaks = seq(from = 4000,to = 13000,by = 1000), limits = c(4000,13000))
  +
  theme(axis.text.x=element_text(angle=0, hjust=1)) + geom_line(aes(y = d1,x = d2),
    linewidth = 0.4, color = "#bf0d1c") +
  theme_minimal()

MAE <- mean(abs(out_sample - prev$mean))
MAE

# MODELO ARIMA - TRANSFORMAÇÃO DE BOX-COX

lambda_auto = serie_temp %>% BoxCox.lambda() lambda_auto

```



```

serie_temp_boxcox = serie_temp %>% BoxCox(lambda_auto) serie_temp_boxcox

serie_temp_boxcox %>% plot() serie_temp %>%
plot()

tseries::kpss.test(serie_temp_boxcox) mstl(serie_temp_boxcox, s.window =
"periodic") %>% plot()

# Gráfico da Série serie_temp_boxcox %>% autoplot() + geom_line(color =
"#1f0866", linewidth = 0.8) + xlab("Anos") + ylab("Depósitos Totais (bilhões de
dólares)") + scale_x_continuous(breaks = seq(from = 1983,to = 1993,by = 1), limits
= c(1983,1993)) +
  scale_y_continuous(breaks = seq(from = 3.6,to = 3.71,by = 0.01), limits = c(3.6,3.71)) +
  theme(axis.text.x=element_text(angle=0, hjust=1)) + theme_minimal()

# Operador de diferença (tendência e sazonalidade) ndiffs(serie_temp_boxcox) # 1
diferença      d = 1 diff(serie_temp_boxcox) %>% nsdiffs() # 0 diferença D = 0

# SARIMA(p,1,q)x(P,0,Q) x <- diff(serie_temp_boxcox) # Série
estacionária par(mfrow=c(1,3)) plot(x);acf(x,lag.max = 12*5);pacf(x,
lag.max = 12*5) tseries::kpss.test(x)

# VERIFICANDO MODELOS ADEQUADOS VIA CRITÉRIO AIC CORRIGIDO
melhor_AICc = Inf for(P in 0:2){
  for(Q in 0:2){
    for(p in 0:3){
      for(q in 0:3){ fit = Arima(serie_temp, order = c(p,1,q), seasonal = c(P,0,Q), lambda
        = "auto")
        if(fit$AICC < melhor_AICc){ melhor_AICc = fit$AICC
          cat("p =",p,"",q =",q,"P =",P,"",Q =",Q,
            ",AICc =",fit$AICC, "\n") }
        }
      }
    }
  }
}

# ANÁLISE DE RESÍDUOS DO MODELO 2 ARIMA

```



```

# Modelo SARIMA(1,1,2)x(0,0,2) fit = Arima(serie_temp, order=c(1,1,2),
seasonal=c(0,0,2), lambda = "auto")
fit

# Resíduos
par(mfrow=c(1,2)) E_1 <-
fit$residuals plot(E_1)
E <- fit$residuals %>% window(start=c(1983,2)) plot(E)

# gráficos
par(mfrow=c(1,3)) plot(E)
qqnorm(E); qqline(E) acf(E,
lag.max = 12*5) hist(E)

# verificando estacionariedade tseries::kpss.test(E)

# verificando independência
Box.test(E,lag = 15, type = "Ljung-Box")
Box.test(E,lag = 20, type = "Ljung-Box")

# Normalidade dos resíduos shapiro.test(E)

# Horizonte de previsão

par(mfrow=c(1,1)) prev = forecast(fit, horizonte, level =
95) prev

library(zoo) d1 = out_sample %>%
as.numeric() d2 =
as.yearmon(time(out_sample))

prev %>% autoplot() + geom_line(linewidth = 0.8) +
  xlab("Anos") +
  ylab("Depósitos Totais (bilhões de dólares)") +
  scale_x_continuous(breaks = seq(from = 1983,to = 1995,by = 1), limits = c(1983,1995))
  +
  scale_y_continuous(breaks = seq(from = 4000,to = 13000,by = 1000), limits = c(4000,13000))
  +
  theme(axis.text.x=element_text(angle=0, hjust=1)) + geom_line(aes(y = d1,x = d2),
    linewidth = 0.4, color = "#bf0d1c") +
  theme_minimal()

MAE <- mean(abs(out_sample - prev$mean))

```



```
##### Séries Arima #####
Arima_p = Arima(serie_temp, order=c(1,1,3), seasonal=c(1,0,1))
Arima_box_cox = Arima(serie_temp, order=c(1,1,2), seasonal=c(0,0,2),
lambda = "auto")

##### Séries ETS #####
ets_p <- ets(serie_temp, model =
"MAM")
ets_boxcox <- ets(serie_temp, model = "AAA", lambda = -
.24)

#####

## Estudo de desempenho preditivo por janela deslizante;
#####

# Tamanho da série
n =
length(serie_temp)

#####

### Funções para a função tsCV
f_arima =
function(x,h){
  forecast(Arima(x,order = c(1,1,3),
seasonal=c(1,0,1)),h=h) } cv_arima =
tsCV(serie_temp,f_arima, h=5,initial=n-14)
#####
f_arima_box_cox =
function(x,h){
  forecast(Arima(x,order=c(1,1,2), seasonal=c(0,0,2), lambda =
'auto'),h=h)
} cv_arima_box_cox = tsCV(serie_temp,f_arima_box_cox,
h=5,initial=n-14)
#####
f_ets =
function(x,h){

  forecast(ets(serie_temp, model = "MAM"),h=h)
```



```

} cv_ets = tsCV(serie_temp,f_ets,
h=5,initial=n-14)

#####
f_ets_box_cox =
function(x,h){
  forecast(ets(serie_temp, model = "AAA", lambda = -
.24),h=h)
} cv_ets_box_cox = tsCV(serie_temp,f_ets_box_cox,
h=5,initial=n-14)
#####

## Cálculo do erro absoluto médio para cada horizonte de previsão
#####
mae_arima = cv_arima %>% abs() %>%
colMeans(na.rm=T)
mae_arima_box_cox = cv_arima_box_cox %>% abs() %>%
colMeans(na.rm=T)
mae_ets = cv_ets %>% abs() %>%
colMeans(na.rm=T)
mae_ets_box_cox = cv_ets_box_cox %>% abs() %>%
colMeans(na.rm=T)
tabela_p = cbind(as.numeric(mae_arima),
as.numeric(mae_ets))
tabela_box = cbind(mae_arima_box_cox,
mae_ets_box_cox)
tabela_mae = data.frame(mae_arima, mae_ets, mae_arima_box_cox,
mae_ets_box_cox)

knitr::kable(tabela_mae)

### ERRO ABSOLUTO MÉDIO - MAE
colnames(tabela_mae) <- c('ARIMA',
'ETS',
'ARIMA Transformada', 'ETS Transformada')

##### Gráficos de Linhas
tf_tidy_dt =
function(dt){
  colnames(dt) =
c('Arima','ETS')

  data = pivot_longer(as.data.frame(dt),cols = c('Arima','ETS')) %>%

```

```

mutate(rep(1:5,each = 2))
      colnames(data) =
c('Modelo','MAE','ID')

return(data)
}

#tabela_p = tf_tidy_dt(tabela_p)

#tabela_box = tf_tidy_dt(tabela_box) #mudar nome

tabela_total = cbind(as.numeric(mae_arima),
as.numeric(mae_ets),mae_arima_box_cox,mae_ets_box_cox)
colnames(tabela_total) = c('Arima','ETS','Arima Transformado','ETS
transformado')
data = pivot_longer(as.data.frame(tabela_total),cols =
c('Arima','ETS','Arima Transformado','ETS transformado')) %>%
mutate(rep(1:5,each = 4))
colnames(data) =
c('Modelo','MAE','ID')

ggplot(data, aes (x = ID,y = MAE, group = Modelo, colour = Modelo))+
geom_line(size = 1) +geom_point(size = 2)+ labs(x = 'h', y =
'MAE')+ theme_bw()+ theme(axis.title.y
=element_text(colour='black' , size=14), axis.title.x
=element_text(colour='black' , size=14), axis.text
=element_text(colour='black',size=13.5), panel.border =
element_blank(),
plot.title = element_text(size = 24, face = "bold"),
axis.line =element_line(colour='black'))+
ggtitle('Erros de previsão para cada horizonte dos modelos selecionados')+
scale_colour_manual(values = c("#009999", "#B2182B", "#0000FF", "#006f33"))

```

```

pacman::p_load("purrr", "glue", "kableExtra")

format_tab <- function(df, caption, ...)
{
  tabela <- kable(df, caption =
caption, booktabs = T, ... ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),
full_width = F
  )
  return(tabela)
}

box_cox__ <- function(x, lambda) {
if (lambda == 0) {
return(log(x))
}
return((x^lambda - 1) / lambda)
}

box_cox <- function(x, lambda, path_to_plot)
{
  lambda <- as.vector(lambda)
  n_plots <- length(lambda)
  transforms <- purrr::map(lambda, ~ box_cox__(x, .))
png(path_to_plot)
  par(mfrow = c(3, ceiling(n_plots / 3)))
  for (i in 1:n_plots) {
    plot(transforms[[i]], main = glue::glue("lambda = {lambda[i]}"))
  }
dev.off()
  return(transforms)
}

diff_series <- function(x, lag) {
diffs <- x %>% forecast::ndiffs()
if (diffs) {
  x <- x %>% diff(diffs)
  sdiffs <- x %>% forecast::nsdiffs()
} else {
  sdiffs <- x %>% forecast::nsdiffs()
}
if (sdiffs) {
  x <- x %>% diff(forecast::nsdiffs, lag = lag)
}

return(list("ts" = x, diff_simples = diffs, diff_sasonal = sdiffs))
}

```



```

model_selection_ets <- function(models, train) {
  aics <- models %>%
    purrr::map_dbl(~ forecast::ets(y = train, model = .x)[["aicc"]])
  df_select <- data.frame("ETS" = models, "AICc" = aics) %>%
    dplyr::arrange(desc(AICc)) return(df_select)
}

residuals_analysis <- function(model, path_to_plot) {
  residuals <- model[["residuals"]]

  png(path_to_plot)
  par(mfrow = c(2, 2))
  plot(residuals)
  qqnorm(residuals)
  qqline(residuals)
  acf(residuals)
  pacf(residuals) dev.off()

  trend_test <- tseries::kpss.test(residuals, null = "Trend")
  independence_test <- Box.test(residuals, lag = 15, type = "Ljung-Box")
  normality_test <- shapiro.test(residuals) testes_df <- data.frame(
    "Teste" = c("KPSS", "Ljung-Box", "Shapiro-Wilk"),
    "Estatística do Teste" = c(
      trend_test[["statistic"]],
      independence_test[["statistic"]],
      normality_test[["statistic"]]
    ),
    "p-valor" = c(
      trend_test[["p.value"]],
      independence_test[["p.value"]],
      normality_test[["p.value"]]
    )
  )
  return(testes_df)
}

prediction_plot <- function(forecast_prediction) {
  return(autoplot(forecast_prediction) +
    xlab("Anos") +
    ylab("Depósitos Totais (bilhões de dólares)") +
    scale_x_continuous(
      breaks = seq(from = 1983, to = 1995, by = 1),
      limits = c(1983, 1995)
    ) +
    geom_line(linewidth = 1.3) +
    theme_bw())
}

```



```

}

MAE <- function(train, test, model, h, name) {
  if (name %in% c("ses", "holt")) {
    predictions <- model(train)
  } else if (name %in% c(
    "ARIMA (manual)", "ARIMA (manual)(Box-Cox)",
    "ETS (manual)", "ETS (manual)(Box-Cox)"
  )) {
    predictions <- model %>% forecast::forecast(h = h)
  } else {
    predictions <- model(train) %>% forecast::forecast(h = h)
  }
  return(mean(abs(test - predictions[["mean"]])))
}

if (!require("pacman")) {
  install.packages("pacman")
}
pacman::p_load("forecast", "Mcomp", "tseries", "ggplot2", "ggpubr")
source("auxiliar_functions.r", encoding = "UTF-8")

LAG_MAP <- c("MONTHLY" = 12)

ids <- c(2119, 2539) id_escolhido
<- ids[2]
serie_escolhida <- M3[[id_escolhido]]
serie_escolhida %>% plot()
treinamento <- serie_escolhida$x
teste <- serie_escolhida$xx
lag <- LAG_MAP[[serie_escolhida$period]] horizonte
<- serie_escolhida$h

# Item a) mstl(treinamento)
%>% plot()

# Item b)
kpss.test(treinamento, null = "Trend")
out_diff_trein <- treinamento %>% diff_series(lag = lag)
c(out_diff_trein$diff_simples, out_diff_trein$diff_sazonal)
treinamento_estacionario <- out_diff_trein$ts

treinamento_estacionario %>%
autoplot() +
  geom_line(linewidth = 1) +
  theme_bw()

```

```
par(mfrow = c(1, 2))  
acf(treinamento_estacionario)
```



```

pacf(treinamento_estacionario)

box_cox(treinamento, lambda
= seq(-3, 3, 1)[-5],
  path_to_plot = "assets/box_cox_escolha.png"
)
treinamento_transformado <- box_cox__(treinamento, lambda = 0)
treinamento_transformado %>% kpss.test(treinamento, null = "Trend")
out_diff_trein_trans <- treinamento_transformado %>% diff_series(lag =
lag)
c(out_diff_trein_trans$diff_simples, out_diff_trein_trans$diff_sazonal)
trein_trans_estac <- out_diff_trein_trans$ts

par(mfrow = c(1, 2))
acf(trein_trans_estac)
pacf(trein_trans_estac)

# A FAZER: Escolher os modelos baseados nos gráficos ou rodar Loops para
valores pequenos de p, q, P, Q

# Item c)
ets_models <- c("AAA", "MAA", "MAM")
model_selection_ets(ets_models, treinamento) %>%
format_tab("", format = "latex", digits = 3) ets_mod
<- ets(treinamento, model = "MAM")
ets_mod_boxcox <- ets(treinamento, model = "AAA", lambda = -.24)
min(box_cox__(treinamento, lambda = -0.24))
# Item d)

residuals_analysis(ets_mod, "assets/ets_mod.png") %>% format_tab("",
digits = 3, "latex")
residuals_analysis(ets_mod_boxcox, "assets/ets_mod_boxcox.png") %>%
format_tab("", digits = 3, "latex") # A FAZER: Incluir os modelos
ARIMA

# Item e)

# item f)

df_teste_plot <- data.frame(
"x" = as.vector(time(teste)),
"y" = as.vector(teste)
)
# Tabelas nas próximas duas linhas
(pred_ets <- forecast(ets_mod, h = horizonte, level = 95, bootstrap =
TRUE))
(pred_ets_boxcox <- forecast(ets_mod_boxcox, h = horizonte, level = 95))
p3 <- pred_ets %>% prediction_plot() +

```

```

    geom_line(
      data = df_teste_plot, aes(x = `x`, y = `y`),
      color = "red", linewidth = 0.5
    ) +
    scale_y_continuous(breaks = seq(4000, 16000, 4000)) +
    coord_cartesian(ylim = c(3800, 16000))

p4 <- pred_ets_boxcox %>%
prediction_plot() + geom_line(
  data = df_teste_plot, aes(x = `x`, y = `y`),
  color = "red", linewidth = 0.5 )

ggarrange(p3, p4) %>%
  ggsave(filename = "assets/predictions.png", .)

# item g) arima_mod
<- Arima(
  y = treinamento, order = c(1, 1, 3), seasonal = c(1, 0, 1)
)
arima_mod_boxcox <- Arima(
  y = treinamento, order = c(1, 1, 2), seasonal = c(0, 0, 2), lambda = 0.24
)
models <- list(
  "ARIMA (manual)" = arima_mod,
  "ARIMA (manual)(Box-Cox)" = arima_mod_boxcox,
  "ETS (manual)" = ets_mod,
  "ETS (manual)(Box-Cox)" = ets_mod_boxcox,
  "auto.arima" = auto.arima,
  "ses" = ses, "holt" = holt,
  "ets" = ets, "stlf" = stlf,
  "bats" = bats, "tbats" = tbats
)
MAEs <- map2_dbl(models, names(models), ~ MAE(
  model = .x, train = treinamento, test = teste, h = horizonte, name = .y
))
data.frame("Implementação" = names(MAEs), "MAE" = unname(MAEs)) %>%
format_tab(
  glue("Erro médio absoluto para as previsões de {horizonte} passos a
frente."),
  digits = 2, format = "latex"
)

```