

Modelos de series temporais aplicadas nos dados de exportação Brasileira

Autor: Oliveira, S.M

Resumo:

A exportação Brasileira tem tido um crescimento acelerado nas ultimas duas décadas, houve um aumento significativo em diversos setores da economia brasileiras, dentre este setores destacam-se as commodities agrícolas, o país já é um dos maiores exportadores de soja, milho e café, neste trabalho objetivou-se avaliar a quantidade em kg de todas as exportação realizadas pelo Brasil a partir do ano de 1999 e a partir de então ajustar um modelo de series temporais para fins de estudo de previsão. Foram avaliados 4 modelos Arima Sazonal de ordem $(0,1,3) \times (2,0,0)_{12}$, para serie original, Arima Sazonal de ordem $(2,1,1)(2,0,0)_{12}$ para série tomada de uma transformação Box-cox, Modelo de alisamento Exponencial ETS(MAM) para serie original e ETS (AAA), para serie tomada de uma transformação de box-cox. Dos quatro modelos ajustados o que obteve a melhor qualidade de ajuste foi o modelo ETS(MAM) sendo o mais indicado para estes dados.

Palavras Chaves: Arima sazonal, Alisamento Exponencial, Box Cox, Previsão, Janela deslizante

Abstract

Brazilian exports have had an accelerated growth in the last two decades, there has been a significant increase in several sectors of the Brazilian economy, among these sectors agricultural commodities stand out, the country is already one of the largest exporters of soy, corn and coffee, in this The objective of this work was to evaluate the quantity in kg of all exports carried out by Brazil from the year 1999 onwards and from then on to adjust a time series model for forecasting purposes. Four Seasonal Arima models of order $(0,1,3) \times (2,0,0)_{12}$ were evaluated for the original series, Seasonal Arima of order $(2,1,1)(2,0,0)_{12}$ for the original series. taken from a Box-cox transformation, Exponential Smoothing Model ETS(MAM) for original series and ETS (AAA), for series taken from a box-cox transformation. Of the four adjusted models, the one that obtained the best quality of fit was the ETS(MAM) model, being the most suitable for these data.

Keywords: Seasonal Arima, Exponential Smoothing, Box Cox, Forecast, Sliding window

Introdução

A proposta de exploração no mercado brasileiro se iniciou desde a época em que era colônia de Portugal, nesta época inicia-se a exploração do ouro e comercialização no mercado internacional, posteriormente a exportação do Pau Brasil, açúcar e café (Contine et al 2012). O Brasil se tornou um dos maiores exportadores de diversas matéria prima nas últimas duas décadas no cenário internacional, principalmente as commodities agrícolas estas de certa forma elevou o percentual de exportação brasileira (Souza e veríssimo 2013), o Brasil se destaca em algumas culturas como soja, milho e café, de modo geral o país está entre os maiores exportadores mundiais no que tange a culturas de produção de grãos (CONAB 2020), partir deste crescimento faz-se necessários modelos de previsões que viabiliza o conhecimento prévio do comportamento do mercado internacional, tais modelos facilitam na tomada de decisões dos setores estratégicos do país. Dentre os modelos de previsões destaca-se as series temporais. Segundo Morettin e Toloi (2006) Series temporais são sequências de observações de uma variável ao longo do tempo e a partir destas observações é possível identificar um padrão através de um modelo e então estender um horizonte de previsão para a variável de interesse.

Neste trabalho considerou-se uma série histórica de exportação Brasileira, e a partir de então descrever com modelo de series temporais um horizonte de previsão para as próximas exportações do País através dos modelos Arimas. Estes modelos são importantes, pois levam em consideração a sazonalidade estocástica dos dados. Quando o período $s=12$, o modelo denominado SARIMA de ordem $(p,d,q) \times (P,D,Q)_{12}$ (MORETTIN & TOLOI, 2004).

Materiais e Métodos

Os dados da série histórica da exportação foram obtidos no site da Receita Federal Brasileira, o conjunto de dados são referente a observações de quilos exportados a partir do ano de 1999 até 2019, porém o conjunto foi particionado para conjunto de treinamento e conjunto de validação, para o treinamento foi considerado as observações de 1999 até o ano de 2018 e para validação foi considerado do ano 2018 a 2019. Os modelos propostos para esta análise foram Arima sazonal com os dados originais, Arima sazonal tomado por uma transformação de box-cox nos dados

A definição matemática do modelo ARIMA sazonal segundo BOX e JENKINS (1974) é representada da seguinte forma:

$\phi(B)\Phi(Bs)W_t = \theta(B)\Theta(Bs)a_t$ com:

$W_t = \nabla s d \nabla d X_t$ e com os seguintes operadores:

- auto-regressivo não sazonal - $\phi(B) = (1 - \alpha_1 B - \dots - \alpha_p B^p)$;
- auto-regressivo sazonal - $\Phi(Bs) = (1 - \phi_s B^s - \dots - \phi_p B^p s^p)$;
- Média móvel não sazonal. $\theta(B) = (1 + \beta_1 B + \dots + \beta_q B^q)$; $\Theta(Bs) = (1 + \theta_s B^s - \dots - \theta_p B^p s^p)$
- Operador média móvel sazonal; ∇d
- Operador diferença não sazonal de ordem d; $\nabla s d$
- Operador diferença sazonal de ordem D.

Para o uso do modelo Arima ser adequado, os resíduos do modelo devem atender as pressuposições de normalidade, independência e estacionaridade, ou seja, os resíduos tem que ser independente com média zero e variância constante. (Morettin 2004)

Também foram utilizados os modelos de alisamento exponencial (ETS) e ETS com a serie tomada por uma transformação de box-cox.

Os modelos ETS em série temporal são modelos que avaliam erro, tendência e sazonalidade. O erro pode ser multiplicativo (M) ou aditivo (A), a tendência pode não ter (N), ser multiplicativa (M), multiplicativa amortecida (Md), aditiva (A) ou aditiva amortecida (Ad) e a sazonalidade pode não ocorrer (N), ser multiplicativa (M) ou aditiva (A). Equações do modelo ETS com erros aditivos e multiplicativos podem ser encontrados em Hyndman et al. (2008).

Para o critério de escolha do melhor modelo foram utilizados a técnica de janela deslizante, o modelo que obter o menor erro absoluto é o escolhido para fazer previsões. O software utilizado para tratamento dos dados foi o software Livre R TEAM 2022.

Resultados

Foi realizado a decomposição da série histórica para verificar o comportamento dos dados, utilizando a função `mstl` do R, esta função decompõe os termos de tendência, sazonalidade e Ruído da série, pode-se identificar na figura 1 os termos de tendência, ou seja, houve um aumento na exportação Brasileira comparando os anos de 1999 a 2018, percebe-se também um padrão sazonal, justificável pois grandes volumes de exportações são referentes ao setor agrário e este setor possui épocas determinadas para colheitas, o que geram ciclos sazonais, e por fim o termo de Ruído da série que aparenta um erro aleatório.

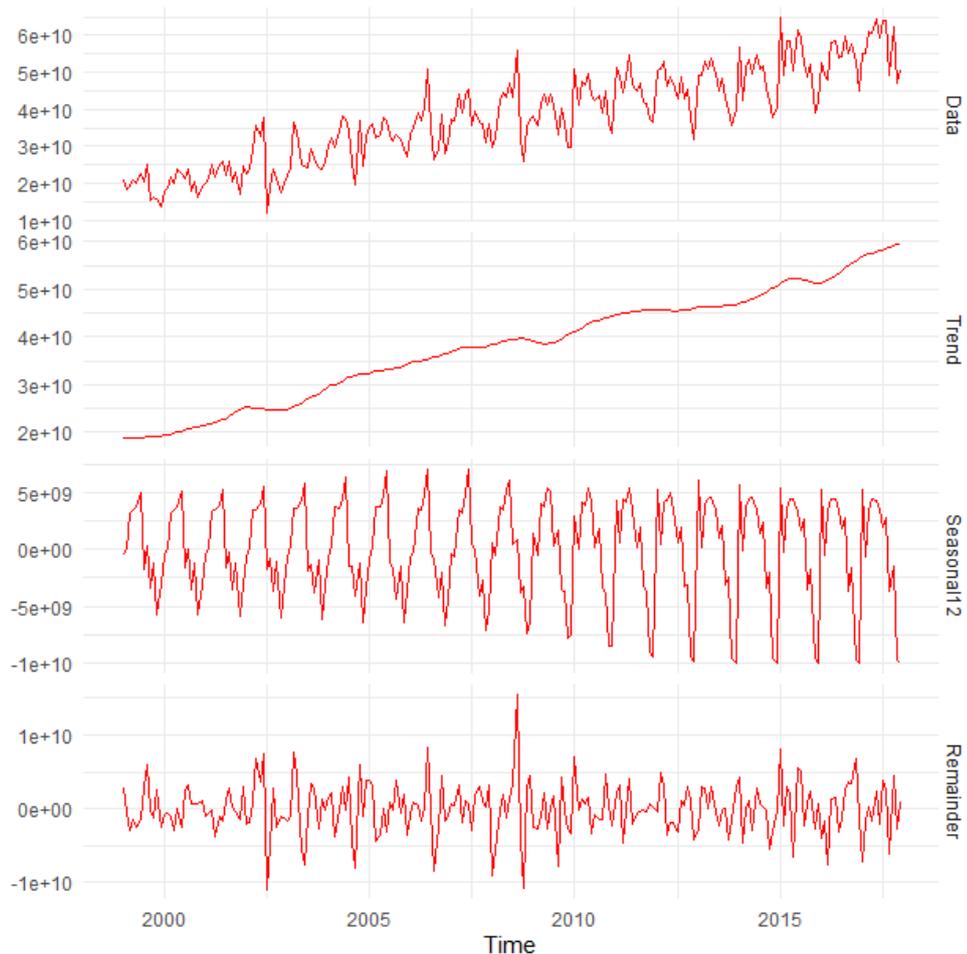


Figura 1: Decomposição da série via *mstl*

Para aplicação dos modelos classes Arima é necessário que a serie seja estacionaria (Moretin 2004) Se uma série $\{Z_t\}$ é não estacionária, logo faz se uma diferença do tipo $W_t = Z_t - Z_{t-1}$, este processo é feito recursivamente até que a serie se torne estacionária, como visto na decomposição na figura 1 os dados apresentam tendência de crescimento e para fazer com que a serie seja estacionária foram necessários aplicar uma diferenciação, esta função de diferença já está implementada no software R basta acionar a função `ndifss()` para checar se necessita de diferença simples e caso necessite aplica a diferença, para checar a sazonalidade utiliza a função `nsdiffs()` esta função retorna a quantidade de diferença sazonal caso necessite para que a serie se torne estacionária. Após este procedimento verificou a necessidade de apenas uma diferença simples e a serie se tornou estacionária.

Na figura 2 tem-se as funções de autocorrelação(ACF) e autocorrelação parcial (PACF), estas funções são as representações gráficas dos coeficientes de autocorrelação em função dos retardos (BOX, JENKINS E REINSEL, 1994), na medida em que este

gráfico demonstram os lags escapam das bandas limites significa que ainda há autocorrelação e o processo não se tornou estacionários, e por estes gráficos em alguns casos pode-se deduzir a ordem do modelo, neste trabalho foi possível deduzir apenas a ordem $d=1$ e $D=0$, não sendo possível obter um modelo completo de forma visual.

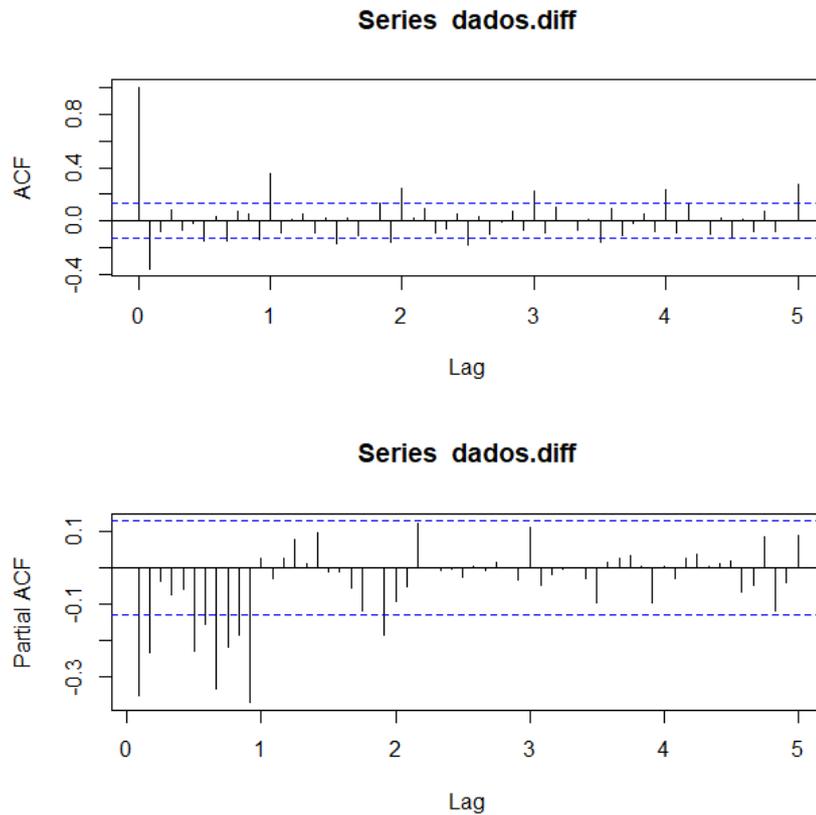


Figura 2: Auto correlação e Autocorrelação parcial

Para escolha do modelo utilizou-se a função `auto.arima` do software com parâmetros $d=1$ e $D=0$, foram testados todas as possíveis combinações de zero a três dos parâmetros p,q,P,Q o modelo que obteve menor AIC corrigido foi selecionado como o melhor modelo para ser ajustado, logo para o modelo 1 obteve-se um $ARIMA(2,1,0)X(1,0,0)_{12}$. Com os coeficientes $AIC=10781.85$ $AICc=10782.23$ $BIC=10802.4$.

Após a seleção do modelo para serie de dados originais realizou-se uma transformação de `boxcox` nos dados e realizou-se novamente o mesmo processo realizado no passo anterior, neste procedimento verificou que para serie se tornar estacionária precisaria tomar uma diferença simples.

Para identificar o modelo Arima na série transformada utilizou-se a abordagem simplista para estimar os parâmetros p,q,P,Q , sendo $d=1$ e $D=0$. Foram testados os valores

de destes parâmetros de 0 a 3 e o modelo com melhor AICc foi escolhido sendo um Arima $(2,1,1) \times (1,0,0)_{12}$. Com $AIC=6691.28$, $AICc=6691.66$ e $BIC=6711.83$.

O terceiro modelo foi proposto como alternativa para os modelos arimas, os modelos de alisamento exponencial, o modelo ETS ajustado para série possui tendência multiplicativa, sazonalidade aditiva e o termo de erro multiplicativo, A figura demonstra a serie decomposta no termo de level, série original e a sazonalidade.

O quarto modelo proposto foi o modelo ETS holt winter, com Erro multiplicativo, Tendência aditiva e sazonalidade Multiplicativa, a figura mostra a decomposição da série em termos de level, sazonalidade e serie observada. O modelo foi gerado através de uma função automática no R que calcula os parâmetros do modelo.

Após definido 4 modelos foram realizadas análise residual, nesta análise é possível determinar se os modelos propostos estão adequados para realizar previsões.

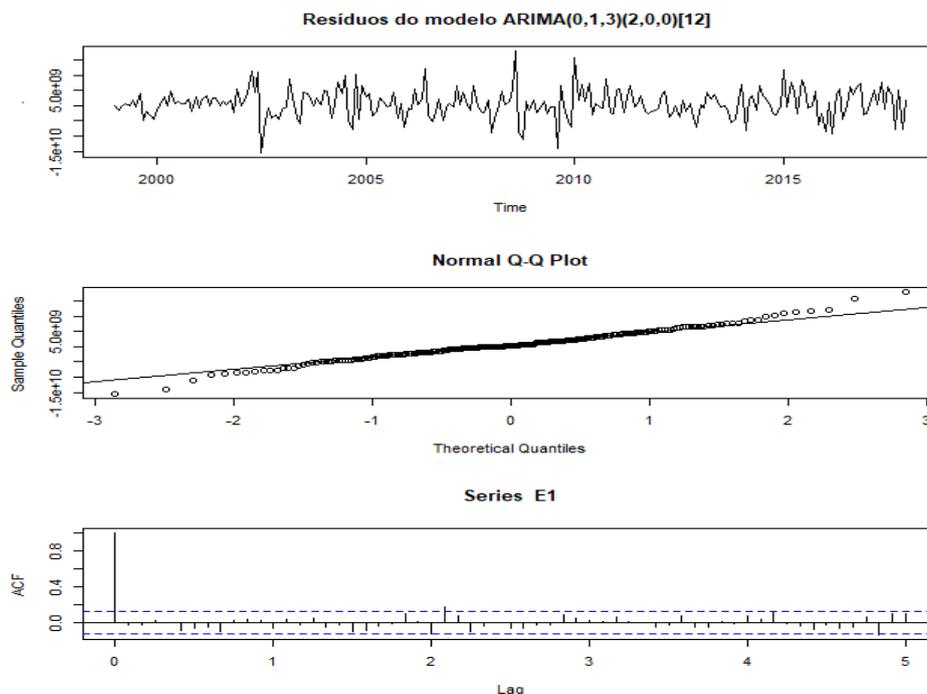


Figura 3: Plot dos resíduos do modelo 1

Na figura 3, o primeiro gráfico demonstra se os resíduos estão de forma aleatória, o segundo gráfico denota a normalidade dos resíduos e o terceiro gráfico demonstra a se o modelo ainda possui autocorrelação, visualmente o modelo está adequado para ser ajustado.

Após avaliação gráfica utilizou-se os testes kpps, que tem por hipótese nula a estacionaridade dos resíduos, utilizou-se também o teste de Ljung Box que testa a independência dos resíduos e foram utilizado o teste de Shapiro wilk para avaliar se os resíduos são normais. Considerando um nível de significância de 1% todos os

pressupostos foram atendidos, se considerar um nível de significância a 5% a normalidade não foi atendida no modelo 1.

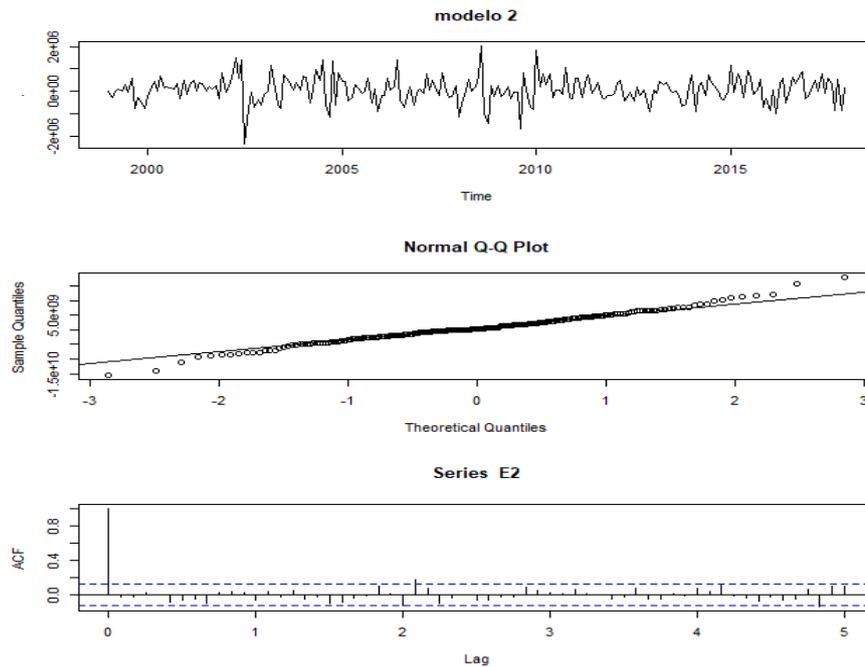


Figura 4: Resíduo do modelo 2

A figura 4 demonstra a análise visual dos resíduos, observado de forma gráfica o modelo aparenta atender os pressupostos, quando submetido através dos testes, os resíduos não atenderam o pressuposto da normalidade a 5% de significância.

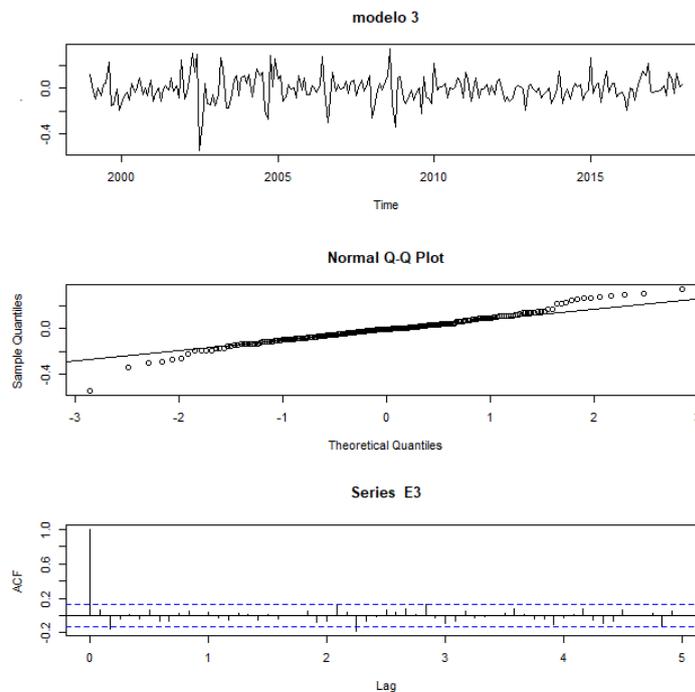


Figura 5: Resíduos do modelo 3

Pelas análises gráficas percebe-se que a normalidade dos resíduos está comprometida, isto corrobora com o teste de Shapiro wilk que aponta a falta de normalidade, já a independência e a estacionaridade foram atendidos.

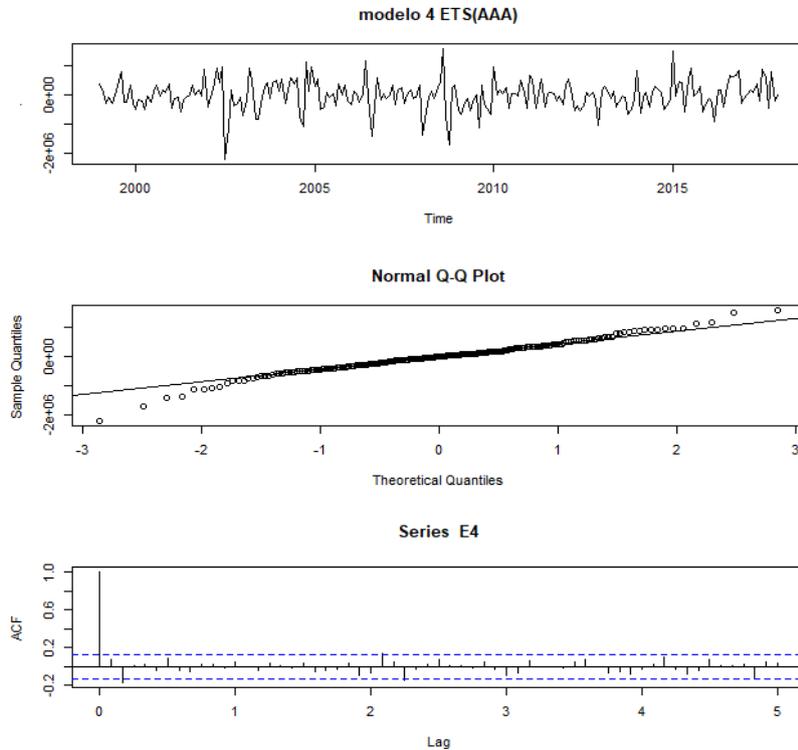


Figura 6: Resíduos do modelo 4

Os resíduos do modelo 4 visualmente são aleatórios e independentes este fato é corroborado com os testes kpss e Ljung Box, nota-se que o gráfico Q-Qplot demonstra alguns pontos fora da diagonal principal, esses pontos podem estar influenciando para os resíduos não serem normais, fato corroborado pelo teste de Shapiro wilk.

Após a avaliação dos resíduos o desempenho do modelo foi observado através da janela deslizante, esta técnica consiste em ajustar o modelo em uma janela de observações e testar na janela complementar, Estudos relacionados à estratégia da janela deslizante podem ser encontrados em Zheng, 2006; Sapucci et al., 2007, foi escolhida as 14 ultimas observações para fazer esta validação e um horizonte de 5 observações para realizar os testes da janela deslizante, os resultados são avaliados pelo menor Erro Absoluto Médio, que estão disposto na figura 7 e 8.

MAE por horizonte de predição.

	ARIMA	ARIMA + Box-Cox	ETS	ETS + Box-Cox
h=1	5231505562	4904219639	2934420319	3222776885
h=2	4123423218	5201099380	3146179653	3190121782
h=3	5205727848	5650853420	3281997090	3361023026
h=4	6182040679	6209889487	3660685166	3742625194
h=5	6096688439	5836823273	3971771728	3901159906

Figura 7: Erros absoluto em 5 horizontes

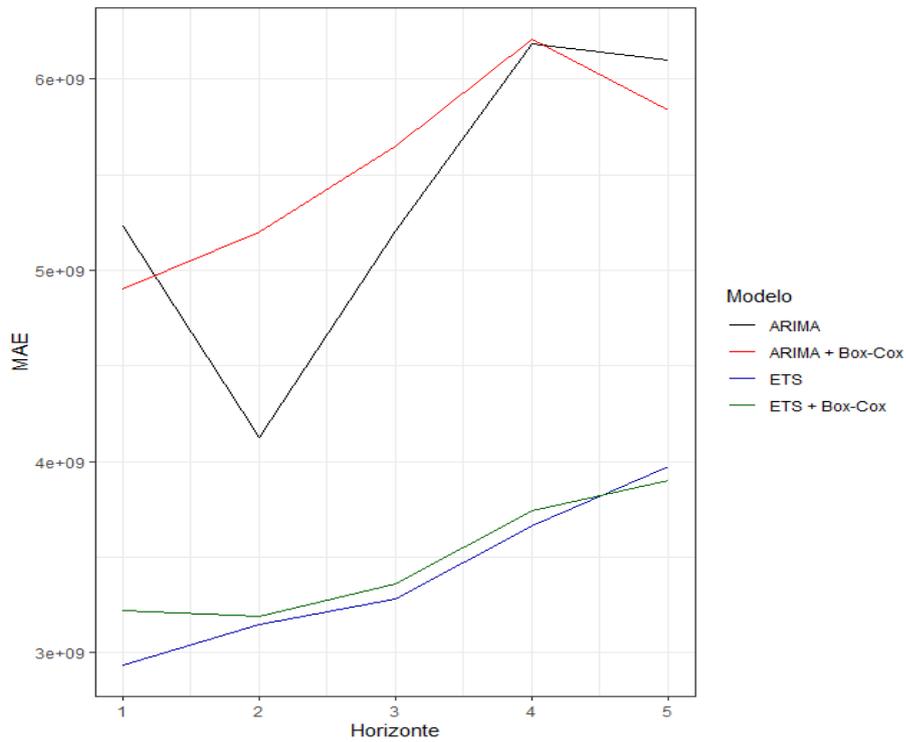


Figura 8: Erros absolutos dos 4 modelos

A figura 8 demonstra que o modelo com menor erro absoluto foi o modelo de alisamento exponencial ETS (MAM), Seguido do Ets (AAA) sendo o modelo Arima sazonal com transformação de box-cox com maior Erro Médio Absoluto.

Foram realizadas previsões pontuais e intervalares os resultados obtidos estão disponibilizados em forma de gráficos na figura 9 e 10 consecutivamente.

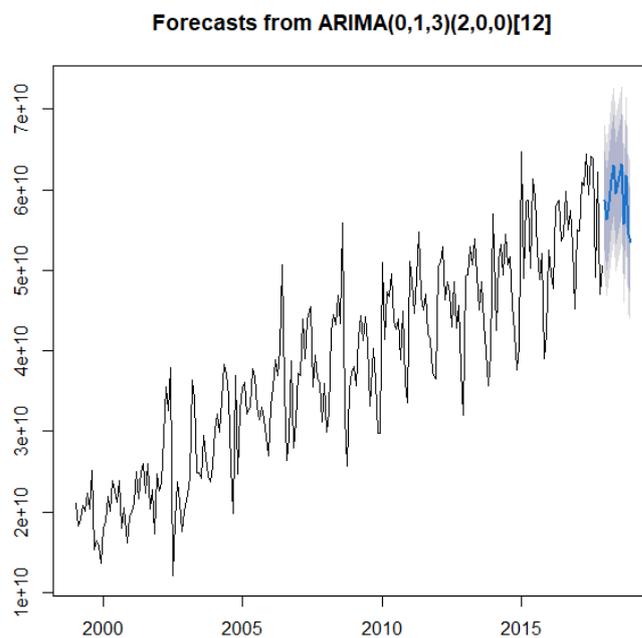


Figura 9: Previsão pontual e intervalar do modelo 1

Na figura 9 tem-se o modelo Arima sazonal $(0,1,3) \times (2,0,0)_{12}$ para os dados sem a transformação de box-cox as partes em cinzas são as bandas intervalares das previsões, e a parte em azul é previsão pontual do modelo, nota-se que o modelo tenta captar a tendência e a sazonalidade da série.

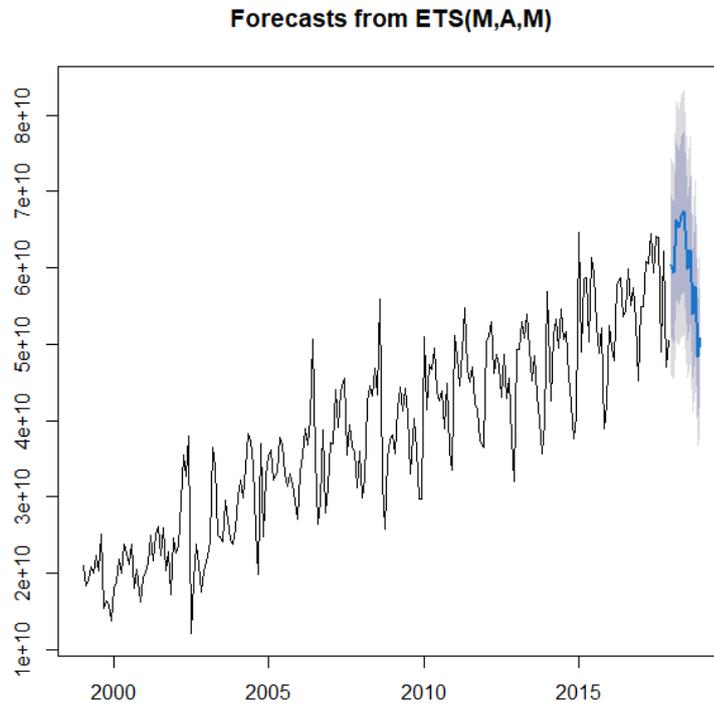


Figura 10: Previsão pontual e intervalar do modelo 3

Nota-se as previsões pontuais em azul e intervalares em cinzas do modelo Ets com erro multiplicativo, tendência aditiva e sazonalidade multiplicativa, este modelo obteve uma boa previsão tanto pontual como intervalar e foi o modelo que na janela deslizante obteve o menor Erro absoluto Médio. Estes resultados corroboram nos dados de teste descritos na tabela abaixo.

MAE nos dados de teste.

	MAE
ARIMA	3606640534
ARIMA + Box-Cox	58786277145
ETS	1902751581
ETS + Box-Cox	2492413045

Figura 11: Métricas dos modelos com os dados de teste

Conclusão

Os modelos Arimas sazonais pode ser ajustado nestes dados o pressuposto da normalidade dos resíduos não foram atendidas, porém este problema pode ser resolvido utilizando bootstrap na previsão. Os modelos de alisamento exponencial demonstraram uma capacidade melhor no ajuste sendo assim mais indicado para avaliação para tratar estes tipos de dados.

Referências Bibliográficas

BOX, G.E., JENKINS, G.M. and REINSEL, G. C. (1994). Time Series: Forecasting and Control (3rd edition). Prentice Hall.

COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB. Acompanhamento da safra brasileira grãos, v. 7 – safra 2019/20 – Décimo primeiro levantamento, Brasília, p. 1-62, agosto 2020. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos>

HYNDMAN, R. J. et al. Previsão com suavização exponencial: A abordagem do espaço de estados. Berlim: Springer-Verlag, 2008

Morettin P. A., Toloí C. M. C. Análise de séries temporais. São Paulo: Edgard Blücher, 2004. 535p

MORETTIN, P. A.; TOLOI, C. M. C. Análise de Séries Temporais. 2. Ed. São Paulo: Blucher, 2006.

SAPUCCI, L. F.; MONICO, J. F. G.; MACHADO, L. A. T.; HERDIES, D. L.; SOUZA; R. A. F. Assimilação do IWV-GPS no Brasil: otimização das estimativas do atraso zenital troposférico em tempo real. Revista Brasileira de Geofísica. vol.25 no.3, São Paulo, 2007.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

ZHENG, Y. Generation of Network-Based Differential Corrections for Regional GNSS Services. Tese de Doutorado, Faculty of the Built Environment and Engineering, Australia, 2006, p. 92.

Arrecadação das Receitas Federais, Conjunto de dados, Disponível em:

<<https://www.gov.br/receitafederal/pt-br/aceso-a-informacao/dados-abertos/receitadata/arrecadacao/serie-historica/arrecadacao-das-receitas-federais-1994-a-2021-1.xlsx/view>>