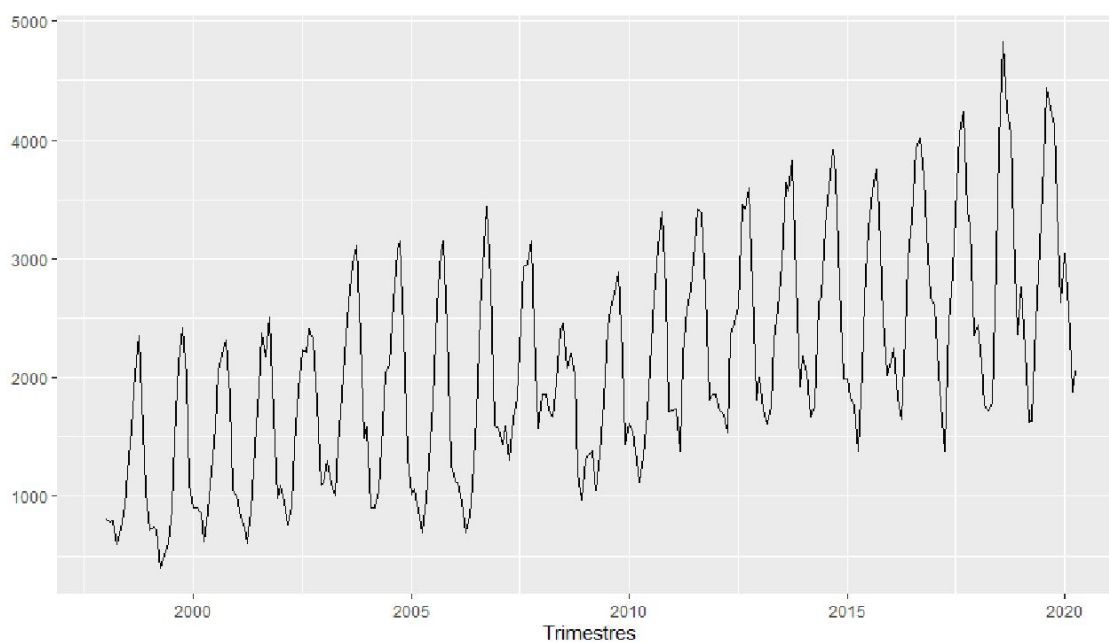


Análise de Séries Temporais

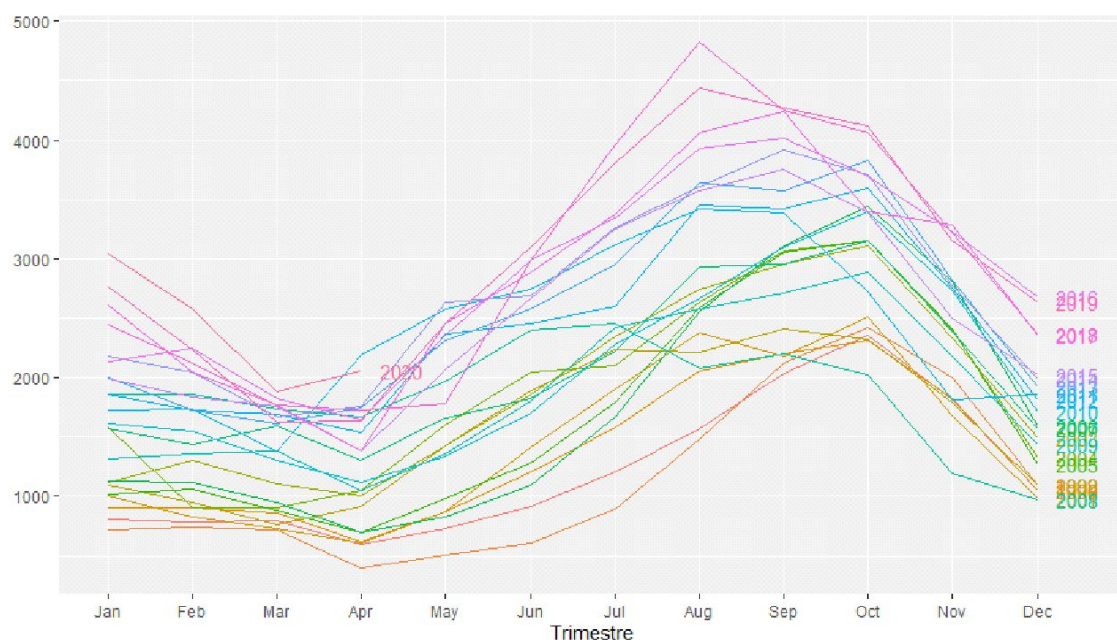
Trabalho Final

Hugo Muniz Albuquerque

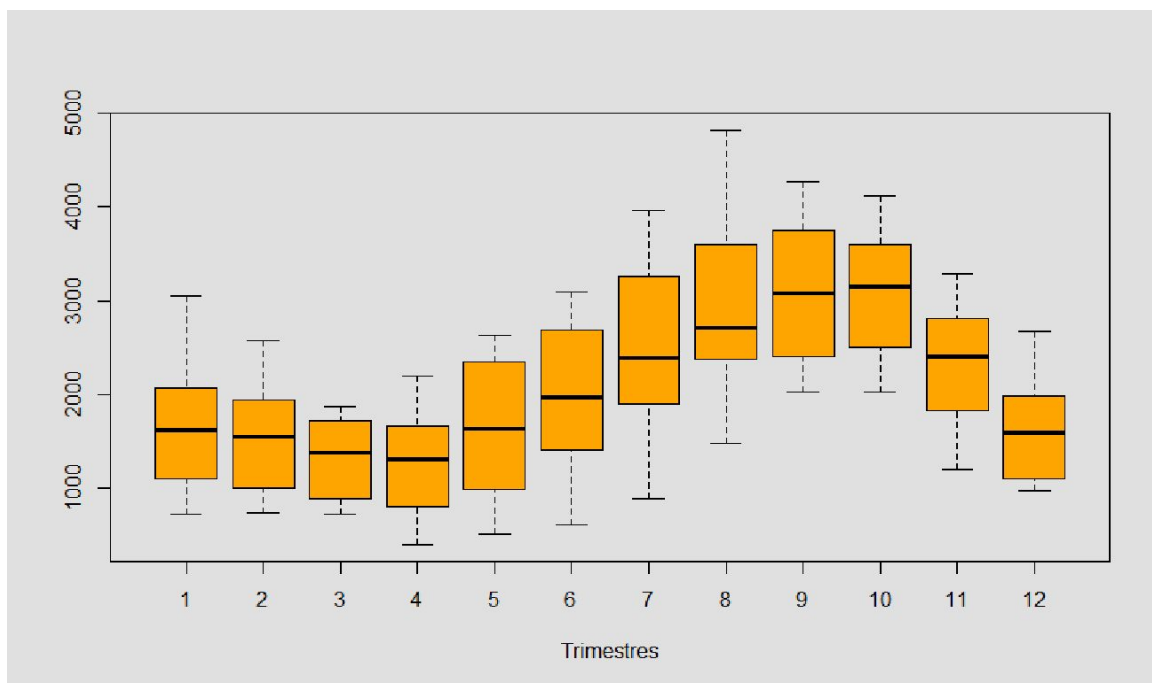
Primeiramente foi realizado a visualização da base inteira, onde pode-se observar uma indicação de uma tendência de crescimento com um padrão de sazonalidade.



Em seguida foi realizado o plot de todos os anos para analisar o padrão de sazonalidade. Como pode-se ver, o consumo de fertilizantes diminui entre janeiro e abril, aumenta entre abril e agosto, e depois volta a cair até o fim de dezembro. Além disso, o nível de consumo cresce com o passar dos anos, visto que em 2020 o consumo atingiu sua máxima.

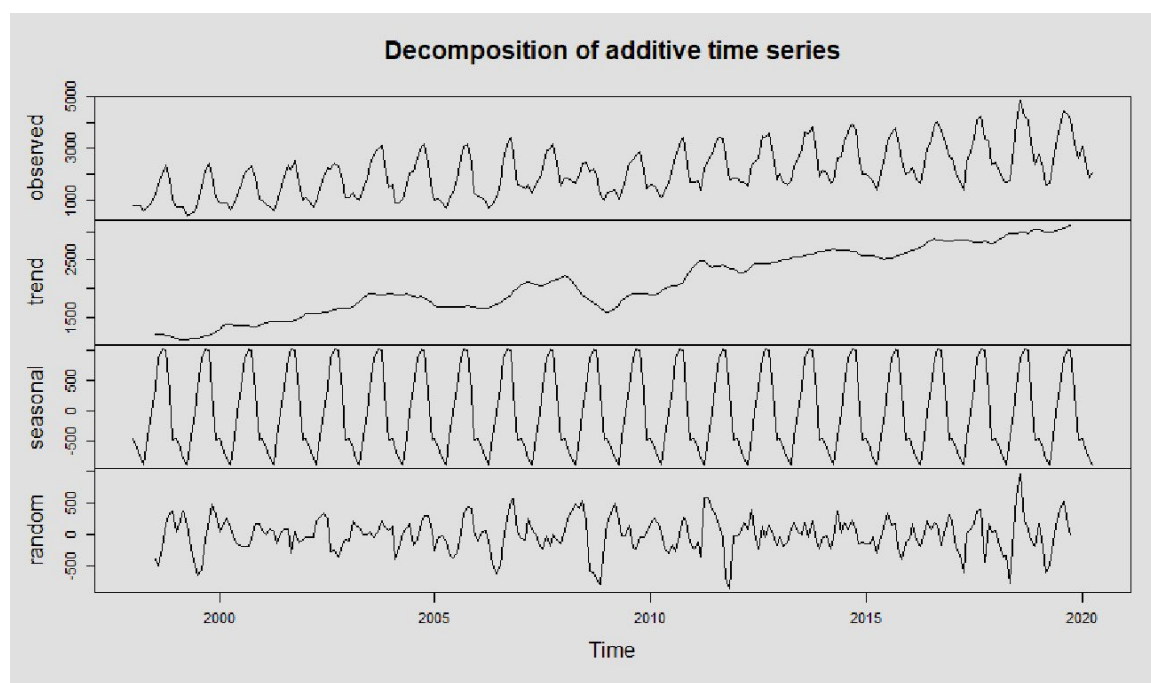


O gráfico boxplot pode confirmar esse movimento de subida e descida do consumo de fertilizantes ao longo dos anos.

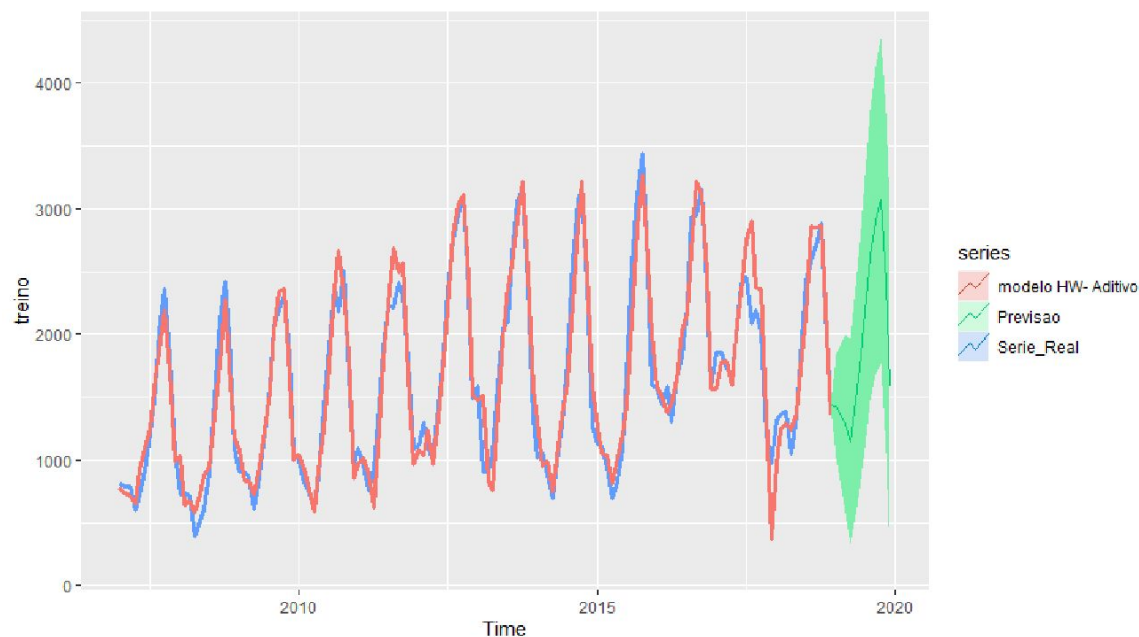


No conjunto de gráficos abaixo, pode-se observar a decomposição da série temporal. Os valores observados apresentam um padrão de crescimento e de sazonalidade como já vistos. A tendência é de crescimento, existe sazonalidade aditiva

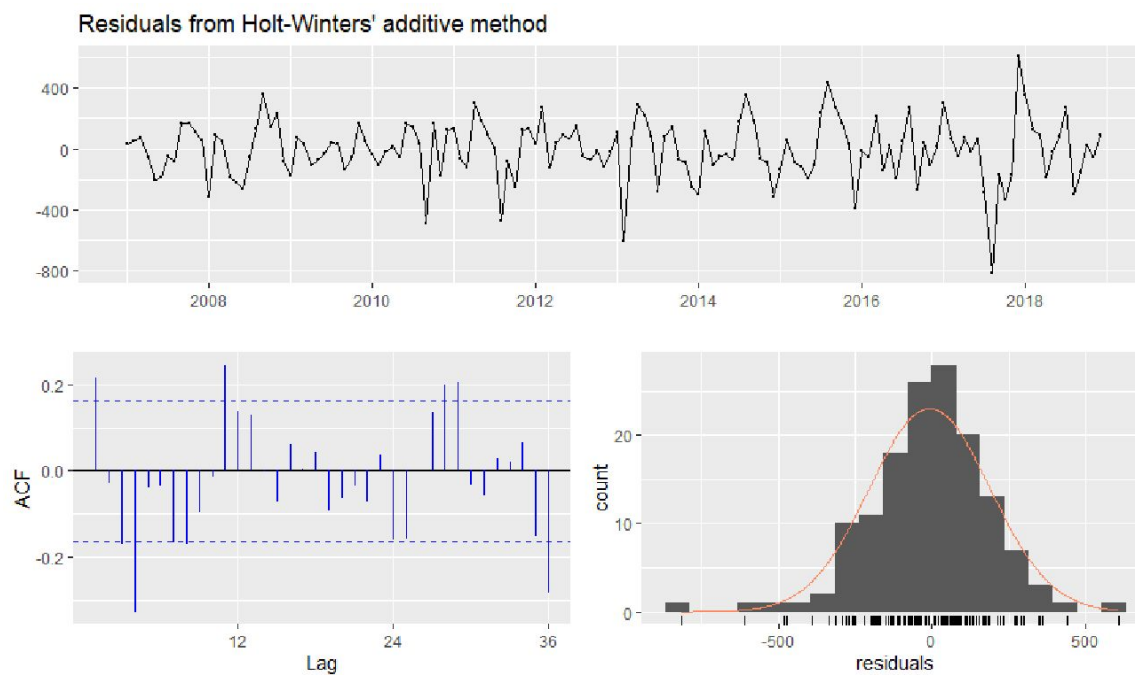
devido ao formato de cone da série sazonalizada e o último gráfico mostra o padrão aleatório da série.



Agora vamos para os modelos. O primeiro modelo obtido foi o Holt Winters com sazonalidade aditiva.

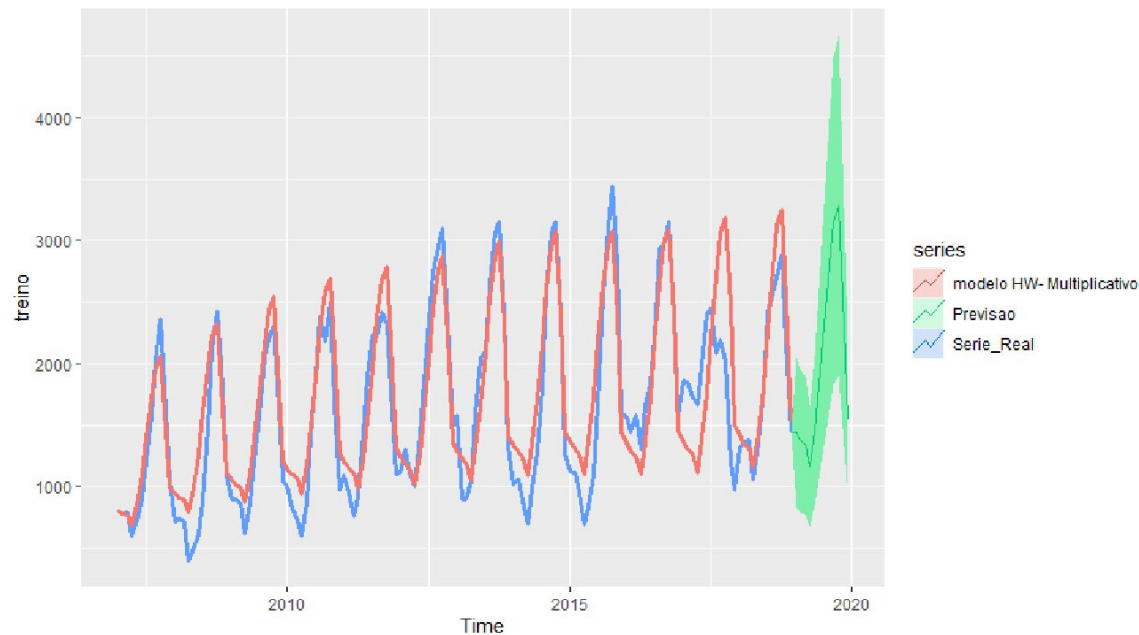


Este modelo aparentou capturar bem os dados e apresentou uma raiz do erro quadrático médio (RMSE) de 196.41 e um AIC de 2270.35.

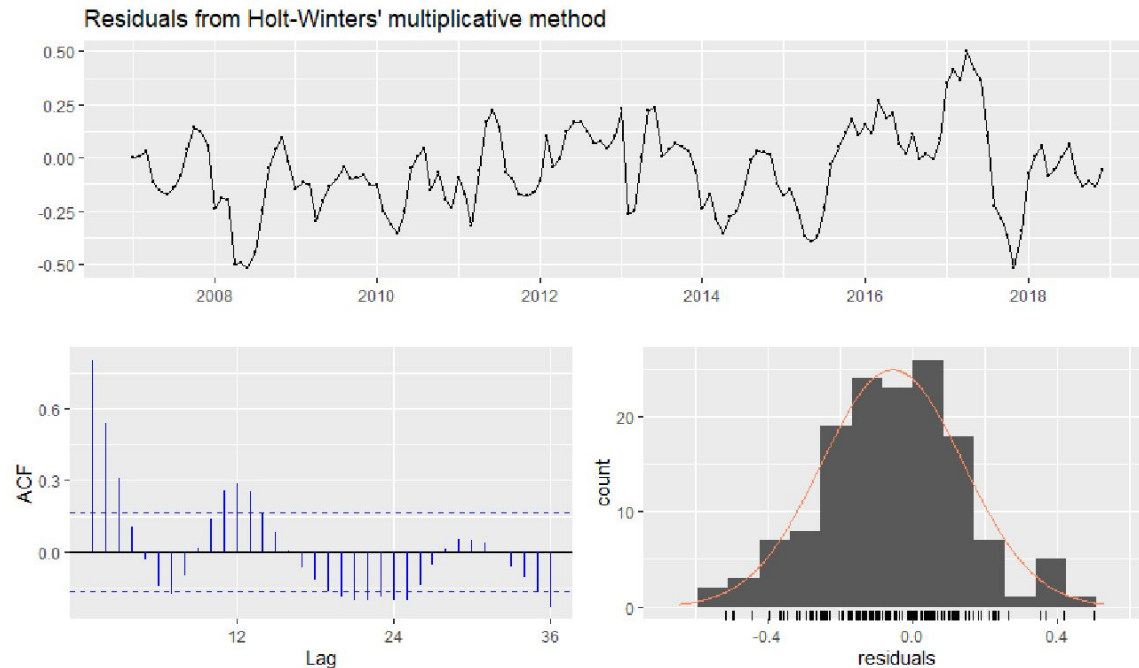


Os resíduos do modelo HW aditivo não parecem ter variância constante no primeiro gráfico, o segundo gráfico abaixo e à esquerda indica que existe autocorrelação serial devido as barras ultrapassando o intervalo de confiança e a distribuição dos resíduos se assemelha a uma normal apesar existirem outliers.

O próximo modelo foi o Holt Winters (HW) com sazonalidade multiplicativa.

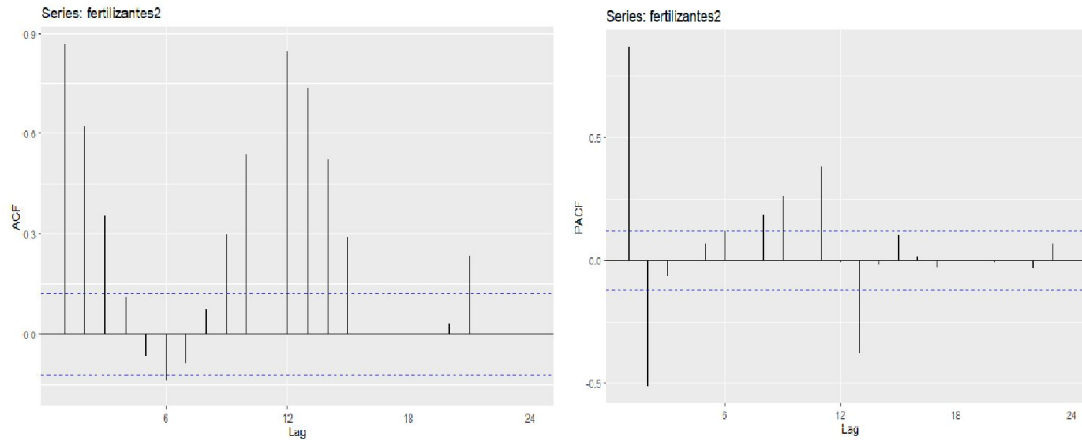


O modelo HW multiplicativo também aparentou capturar bem os dados, porém apresentou um RMSE de 323.78 e um AIC de 2415.665, indicando ser um modelo pior que o HW aditivo.



Os resíduos do modelo HW multiplicativo não parecem ter variância constante no primeiro gráfico indicando possuir mais variância que o modelo aditivo, o segundo gráfico abaixo e à esquerda indica que existe autocorrelação serial devido as barras ultrapassando o intervalo de confiança e a distribuição dos resíduos se assemelha a uma normal apesar existirem outliers.

Por último foi feito o modelo SARIMA, que obteve o menor RMSE e o menor AIC quando aplicado com os seguintes argumentos `stepwise = FALSE`, `seasonal = TRUE` e `lambda = "auto"`. O argumento `lambda = "auto"` aplica uma transformação Box-Cox automaticamente nas variáveis do modelo. O modelo SARIMA obtido possui um RMSE de 189.99 e um AIC de 1226.01, logo foi escolhido o melhor modelo.



O correlograma de autocorrelação serial e de autocorrelação serial parcial mostram visualmente que a série não apresenta padrão estacionário.

Agora vamos para os teste formais de estacionariedade:

Augmented Dickey-Fuller Test

data: fertilizantes2

Dickey-Fuller = -10.174, Lag order = 6, p-value = 0.01

alternative hypothesis: stationary

De acordo com o teste Augmented Dickey-Fuller a série é estacionária, pois o p-valor do teste é menor que o nível de significância de 5% do teste. Logo, rejeita-se a hipótese nula de não estacionariedade. Assim, o teste indica que a série é estacionária com 95% de confiança.

KPSS Test for Level Stationarity

data: fertilizantes2

KPSS Level = 2.644, Truncation lag parameter = 5, p-value = 0.01

O teste KPSS para checar a estacionariedade da série possui como hipótese nula a afirmação de que a série é estacionária (ao contrário do teste ADF e PP). Como

o p-valor obtido foi menor que o nível de significância de 5%, rejeita-se a hipótese nula. Assim, a série não é estacionária com 95% de confiança segundo o teste KPSS.

Phillips-Perron Unit Root Test

data: fertilizantes2

Dickey-Fuller Z(alpha) = -81.213, Truncation lag parameter = 5, p-value = 0.01

alternative hypothesis: stationary

Por último, o teste Phillips-Perron de raiz unitária obteve um p-valor menor que o nível de significância de 5%. Tendo em vista que a hipótese nula é de que a série não é estacionária, ao rejeitar esta hipótese nula o teste indica que a série é estacionária com 95% de confiança.

O modelo obtido com a função `auto.arima` foi o seguinte:

Séries: treino

ARIMA(4,0,2)(2,1,0)[12] with drift

Box Cox transformation: lambda= 0.7067318

Um modelo sarima com $p=4, d=0, q=2$ e $P = 2, D=1$ e $Q=0$ com 12 períodos (meses) e com drift. A seguir foi aplicado a função `t_test` do pacote BETS para verificar a significância dos parâmetros do modelo SARIMA.

	Coeffs	Std.Errors	t	Crit.Values	Rej.H0
ar1	-0.2352293	0.11237869	2.093184	1.977692	TRUE
ar2	0.2133790	0.08091802	2.636978	1.977692	TRUE
ar3	0.5113841	0.08788643	5.818693	1.977692	TRUE
ar4	-0.3935378	0.08504204	4.627567	1.977692	TRUE
ma1	1.3095961	0.08950984	14.630751	1.977692	TRUE
ma2	0.8246969	0.10946644	7.533788	1.977692	TRUE
sar1	-0.4566156	0.09257507	4.932382	1.977692	TRUE
sar2	-0.3751214	0.09674698	3.877345	1.977692	TRUE
drift	0.7022418	0.32930978	2.132466	1.977692	TRUE

Os parâmetros do modelo SARIMA auto.fert são significativos pois o valor das suas estatísticas de teste estão dentro da área de rejeição da hipótese nula, ou seja, são maiores que os valores críticos de aceitação da hipótese nula, logo rejeita-se a hipótese nula de que os parâmetros não são significativos. A seguir vamos verificar os pressupostos para os resíduos do modelo.

Box-Ljung test

data: arima.fert\$residuals

X-squared = 26.867, df = 16, **p-value = 0.04299**

De acordo com o teste Box-Ljung para verificar a ausência de autocorrelação serial, o modelo SARIMA apresentou um p-valor menor que 5% indicando que se deve rejeitar a hipótese nula de que os resíduos do modelo são independentes e identicamente distribuídos (inexistência de autocorrelação serial). O que indica que o modelo possui autocorrelação serial.

ARCH LM-test;

Null hypothesis: no ARCH effects

data: arima.fert\$residuals

Chi-squared = 25.245, df = 36, **p-value = 0.91**

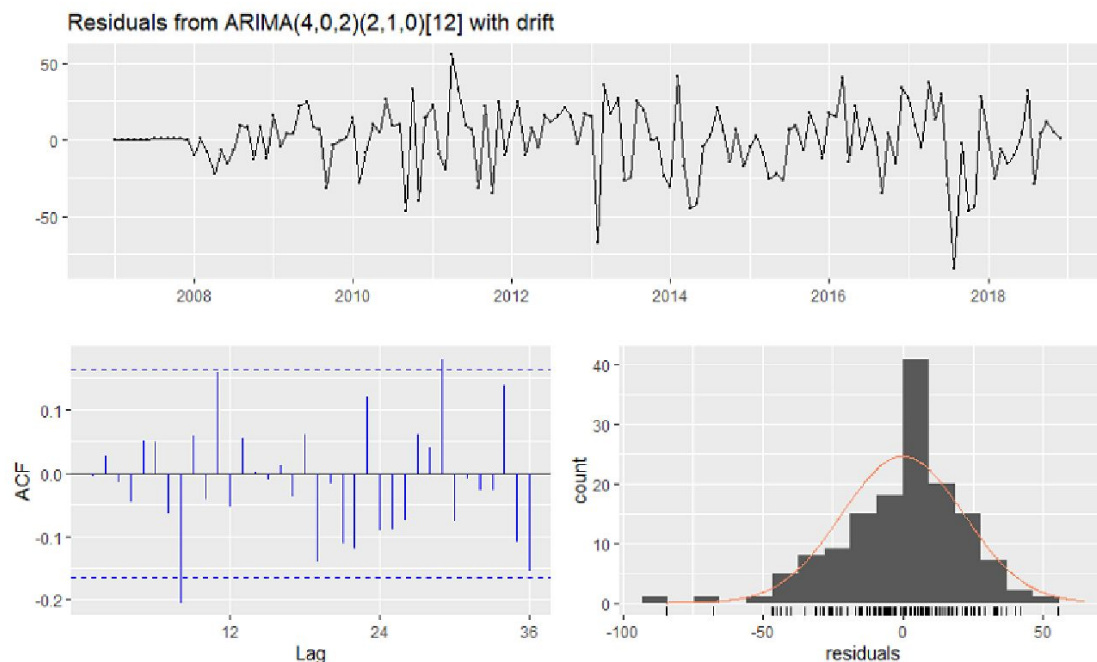
O modelo SARIMA no teste de heterocedasticidade ARCH apresentou p-valor maior que 5% indicando que não deve-se rejeitar a hipótese nula de que a série é homocedástica. Logo, o modelo apresenta variância constante para todas as observações da série.

Jarque-Bera test for normality

data: sarima.3\$residuals

JB = 2.4028, **p-value = 0.1855**

O teste de normalidade dos resíduos Jarque-Bera apresentou p-valor menor que 5% o que indica que deve-se rejeitar a hipótese nula de normalidade dos resíduos. Assim, o teste indica que a distribuição dos resíduos do modelo não é normal. Agora vamos analisar os gráficos dos resíduos do modelo SARIMA.



Os resíduos do modelo SARIMA não parecem ter variância constante no primeiro gráfico (homocedasticidade), pois ao longo dos anos a variação vai aumentando. O segundo gráfico abaixo e à esquerda indica que não existe autocorrelação serial, pois as barras de lag estão dentro do intervalo de confiança do gráfico ACF. O último gráfico apresenta a distribuição dos resíduos do modelo, a qual se assemelha a uma normal apesar de existirem outliers. Assim, o erro e o AIC do modelo SARIMA foram os menores encontrados. Contudo, os pressupostos de ausência de autocorrelação serial e de normalidade dos resíduos não foram atendidos, o que pode gerar previsões enviesadas.

Tabela de dados da base de teste vs ajuste do modelo

	teste	fit	IC95%	IC95%
	2019	2019	lower	upper
Jan	2762	1.703.140	1309.66	2125.39
Feb	2235	1.600.848	1045.82	2219.67
Mar	1624	1.606.283	981.52	2313.11
Apr	1634	1.362.476	743.57	2079.30

May	2458	1.609.937	949.77	2362.41
Jun	3099	1.999.076	1287.48	2794.98
Jul	3811	2.520.453	1751.26	3366.12
Aug	4437	2.773.337	1977.99	3642.44
Sep	4270	2.945.767	2134.46	3829.01
Oct	4116	3.064.105	2242.12	3956.88
Nov	3155	2.289.394	1542.57	3116.27
Dec	2632	1.619.124	955.60	2375.35

Na tabela acima é possível observar que os valores da base teste e dos valores ajustados são semelhantes, apesar de não capturar completamente a variância dos dados. Além disso, a maioria das observações estão dentro do intervalo de 95% confiança. Portanto, podemos concluir que o melhor modelo é o $ARIMA(4,0,2)(2,1,0)[12]$.