Homework 2- Group E

Quarterly data earnings per share of Coca-Cola Company

1. Original data (y) is clearly not stationary, taking consideration of both adf_test result and acf with a decreasing trend. According to arima analysis, there are 1 regular difference and 1 seasonal difference.

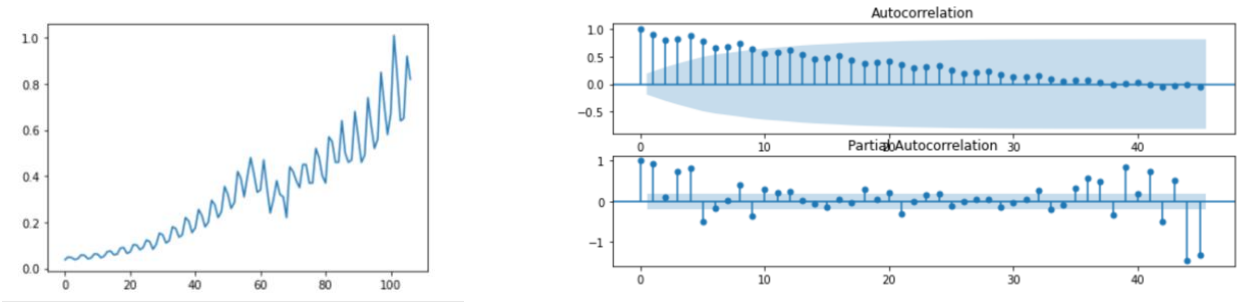

Fig. 1 Original series y

2. To discover a possible model based on suggested differences: Taking **SARIMAX (0,1,0)x(0,1,0) of s=4 model**, the residual becomes stationary around zero mean:

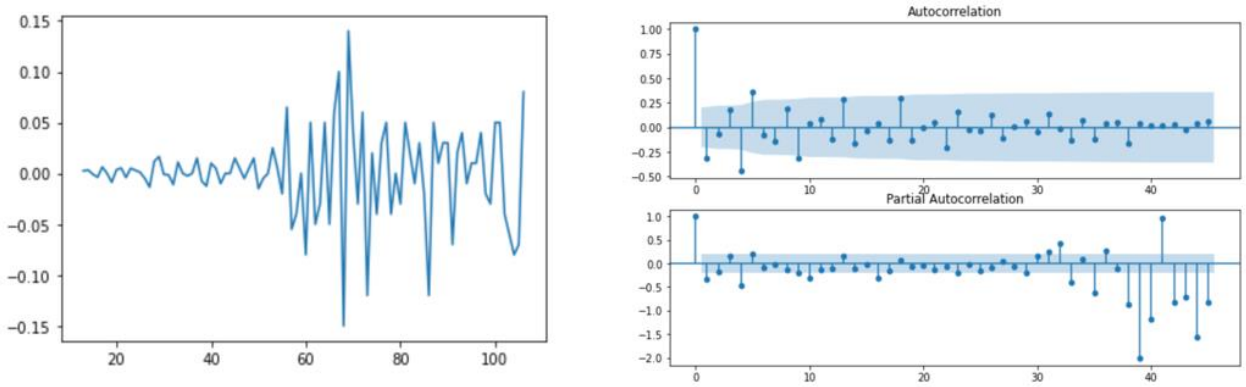

Fig. 2 Residual of SARIMAX (0,1,0)x(0,1,0) of s=4 model

3. Observing the ACF/PACF of residual for SARIMAX (0,1,0)x(0,1,0) of s=4 model: lag 1 and lag 5 of both ACF and PACF are out of bounds, indicating either a AR(1) or MA(1) for the non-seasonal part. Lag 4 for both ACF/PACF are out of bounds, indicating a possible S-MA(1) model for the seasonal part.

SARIMAX Results

Dep. Variable:		value	No. Observations:			107
Model:	SARIMAX(1, 1, 1)x(0, 1, 1, 4)		Log Likelihood			197.680
Date:	Fri, 22 Oct 2021		AIC			-387.359
Time:	19:41:36		BIC			-376.859
Sample:	0		HQIC			-383.107
	- 107					
Covariance Type:	opg					

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.6445	0.082	7.857	0.000	0.484	0.805
ma.L1	-0.9594	0.052	-18.316	0.000	-1.062	-0.857
ma.S.L4	-0.3806	0.105	-3.610	0.000	-0.587	-0.174
sigma2	0.0012	0.000	10.621	0.000	0.001	0.001

4. The parameters are all significant (the absolute value of z is greater than 1.96) for **SARIMAX(1,1,1)x(0,1,1,4) model**. **The residual of this model is WHITE NOISE**, but NOT Gaussian White Noise, NEITHER SWN.

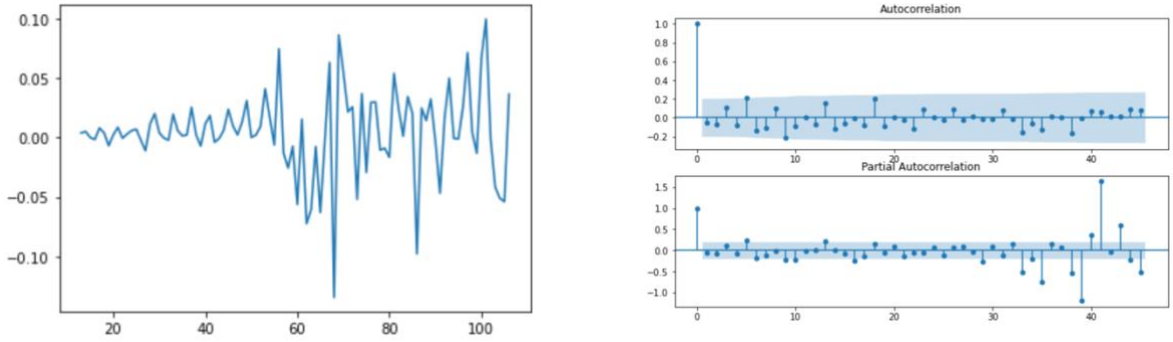

Fig. 3 Residual of SARIMAX (1,1,1)x(0,1,1) of s=4 model

SARIMAX(1,1,1)x(0,1,1,4)	*Statistics (p-value)*	conclusion
Stationary(adf_test)	1.742328416728474e-17	The residual is stationary, and NO transformation is needed.
Linear Model(LjungBox)	0.09483397	Residual is independently distributed; therefore, NO linear model is needed
Normal distribution	3.8048994611017406e-05	The residual is NOT normally distributed
WN	zero mean, constant variance, acf/pacf is zero	The residual of this model is WHITE NOISE
GWN	Residual is not normally distributed	NOT Gaussian WHITE NOISE
SWN	Acf/pacf of Squared residuals have non-zero lags out of bounds	There is a linear relation to explain squared residuals, and the residual is NOT Strict White Noise.

5. To discover a possible model, considering no difference is taken. Observing original y's ACF/PACF, an AR(1) and an S-AR(1) models are estimated. However, the residual only becomes stationary and white noise after taking the 1 seasonal difference.

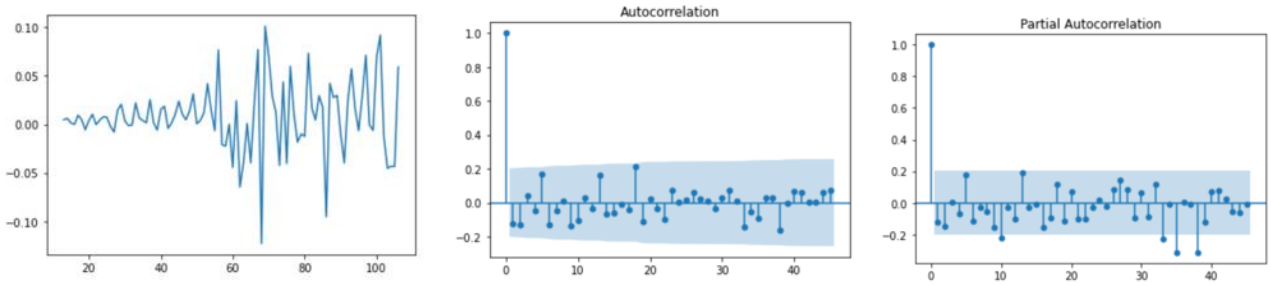

Fig. 4 Residual of SARIMAX (1,0,0)x(1,1,0) of s=4 model

SARIMAX Results

Dep. Variable:	value	No. Observations:	107
Model:	SARIMAX(1, 0, 0)x(1, 1, 0, 4)	Log Likelihood	198.799
Date:	Fri, 22 Oct 2021	AIC	-391.597
Time:	19:18:39	BIC	-383.693
Sample:	0	HQIC	-388.396
	- 107		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.8255	0.063	13.041	0.000	0.701	0.950
ar.S.L4	-0.4360	0.105	-4.153	0.000	-0.642	-0.230
sigma2	0.0012	0.000	10.017	0.000	0.001	0.001

6. The parameters are all significant (the absolute value of z is greater than 1.96) for **SARIMAX(1,0,0)x(1,1,0,4) model**. The residual is WHITE NOISE, but NOT Gaussian White Noise, NEITHER SWN.

SARIMAX(1,0,0)x(1,1,0,4)	*Statistics (p-value/rationale)*	conclusion
Stationary(adf_test)	3.736052668333992e-19 <0.05	The residual is stationary, and NO transformation is needed.
Linear Model(LjungBox)	0.30533436>0.05	Residual is independently distributed; therefore, NO linear model is needed
Normal distribution	0.0002930311602540314 <0.05	The residual is NOT normally distributed
WN	zero mean, constant variance, acf/pacf is zero	The residual of this model is WHITE NOISE
GWN	Residual is not normally distributed	NOT Gaussian WHITE NOISE
SWN	Acf/pacf of Squared residuals have non-zero lags out of bounds	There is a linear relation to explain squared residuals, and the residual is NOT Strict White Noise.

Comparing SARIMAX(1,0,0)x(1,1,0,4) model vs **SARIMAX(1,1,1)x(0,1,1,4) model**

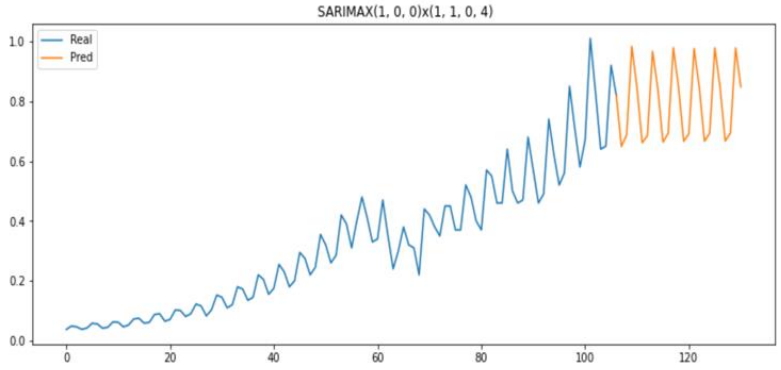

Fig. 5 24 points of predication from SARIMAX (1,0,0)x(1,1,0) of s=4 model.

Variance of prediction is 0.01612

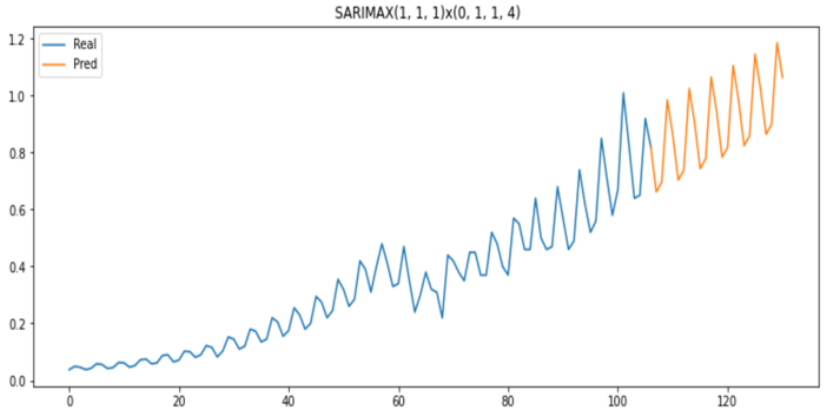

Fig. 6 24 points of predication from SARIMAX (1,1,1)x(0,1,1) of s=4 model.

Variance of prediction is 0.02186

Theoretically, the **SARIMAX(1,1,1) x (0,1,1,4)** model is a better model, judging from greater variance of 24 points of prediction, comparing to the other model. The greater variance indicates better capturing of slope estimation, which is also observable from two graphs above.

However, 24 points of predication cover a period of 6 years. As we try to predict the earning per shares of a multinational beverage company, there are many risk factors that impact the continuing growth of the stock prices within the six years. Therefore, **SARIMAX (1,0,0)x(1,1,0) of s=4 model** might be a more accurate model for a short-term point prediction (less than 4 points prediction that cover only one year).