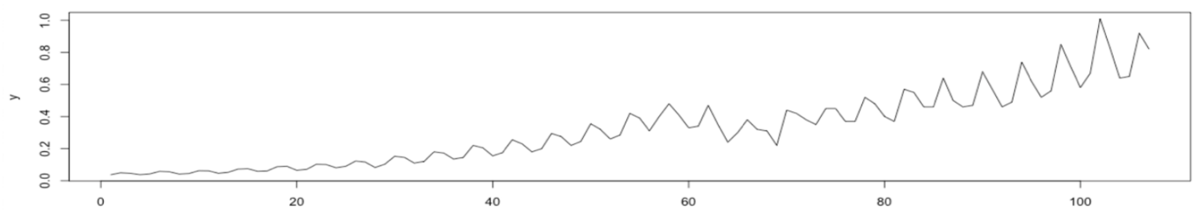
Forecasting Time Series
Homework 2 – Group C

We are using the Box-Jenkins methodology to find at least two linear time-series models, for the quarterly earnings per share of the Coca-Cola Company from the first quarter of 1983 to the third quarter of 2009.

Coca Cola Earnings 1983-2009 (Quarterly data):

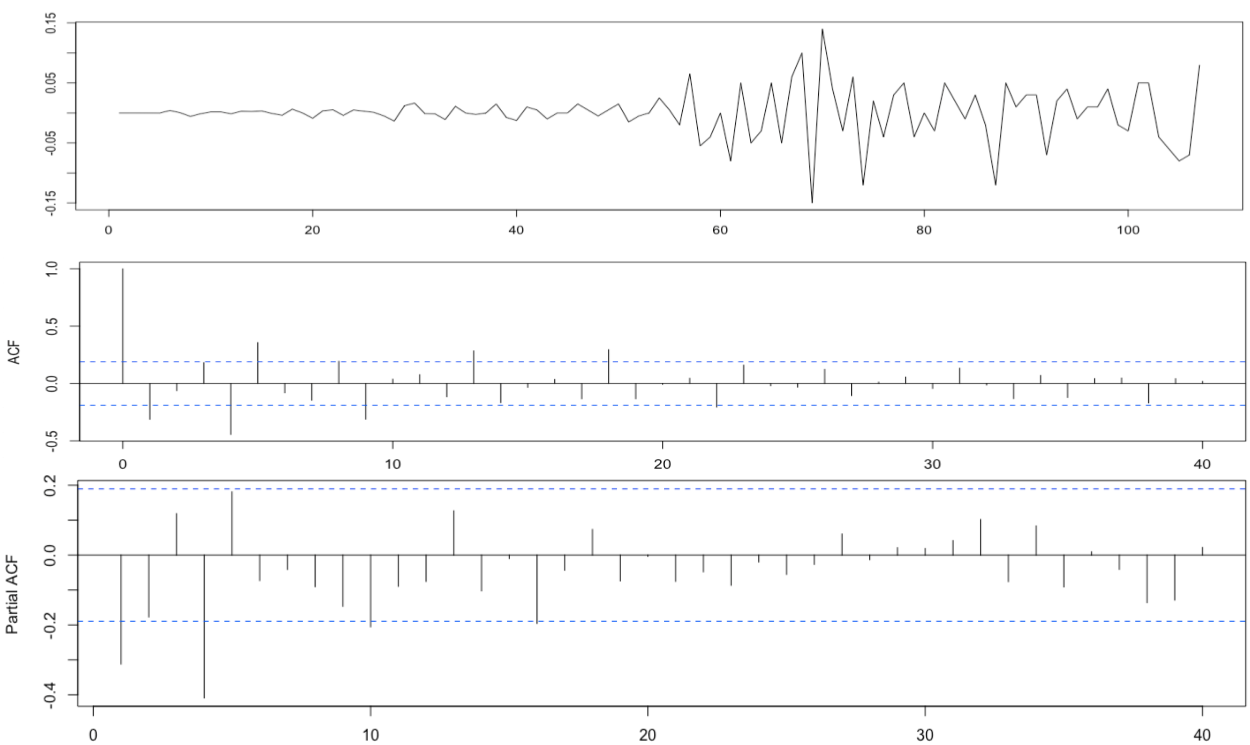


By plotting the earnings data, it was observed that the series (Y_t) is not stationary in the mean but has an increasing trend in the variance that led us to consider the log (Yt) approach as well. Moreover, by plotting the ACF, its sinusoidal shape of the is a strong indication for a cyclical/seasonal trend underlying the data, with many lags out of limits. The Box-Jenkins method is now applied.

Box-Jenkins methodology:

a) **(Y_t)** - S: **4**

1. Seasonal Differences: **1** – Regular Differences: **1**



The series is stationary in the mean and the ACF and PACF have structure to it, so ther to possible models. Thereafter and check if their coefficients were statistically significant.

2. Possible SARIMA **(p, d, q) x (P, D, Q)**, parameter **significance** and **residuals:**

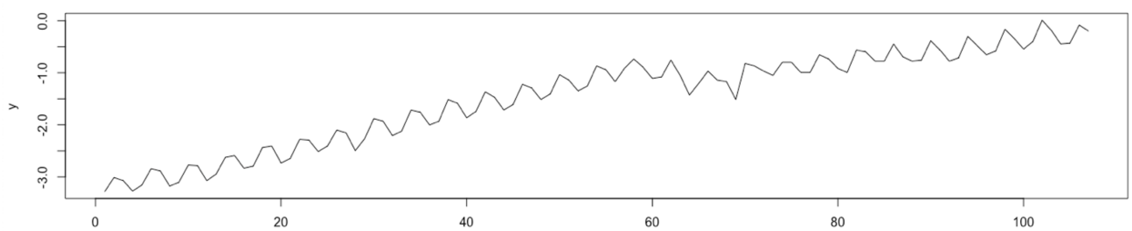
The seasonal and regular lags are examined to explore several autoregressive and moving average models. Although these models were proposed they were subsequently disregarded as their parameters were not statistically significant (e.g. {(0,1,0)x(0,1,1)} or {(0,1,0) x(4,1,0)}).

Thereafter we checked if the residuals for white noise. More proposed models were removed from the list. We then selected two models with a simple structure on the PACF and ACF which allowed us to capture the seasonal trend with a small number of significant parameters and white noise residuals.

SARIMA Model (p,d,q)x(P,D,Q)	Sig. Param	Box Test
(0,1,0)x(1,1,0)	Yes	0.6719
(1,1,0)x(0,1,1)	Yes	0.2497

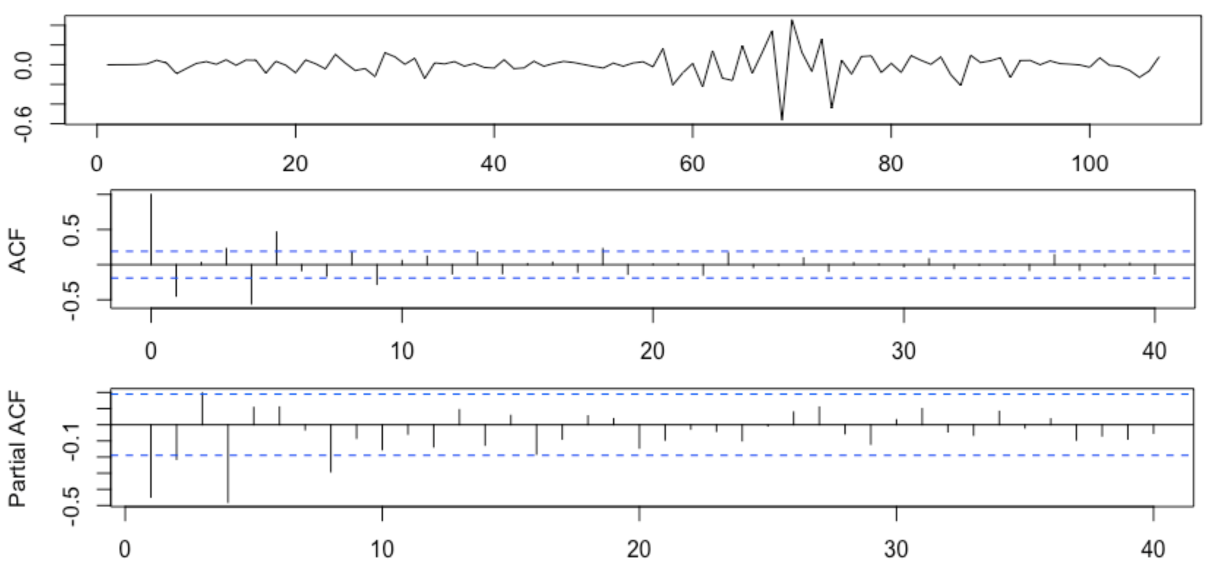
b) **Log (Yt) - S: 4**

From the original time series, we took the **log (Yt)** to ensure that the series is stationary in the variance and proceeded with the Box-Jenkins methodology with the same approach as covered before.



Even though, it is now stationary in the variance, we executed the Dickey Fuller test and applied the necessary regular and seasonal differences to make the series stationary in the mean.

1. Seasonal Differences: **1** – Regular Differences: **1**



2. Possible SARIMA **(p, d, q) x (P, D, Q)**, parameter **significance** and **residuals**:

Log	SARIMA Model (p,d,q)x(P,D,Q)	Sig. Param	Box Test
Yes	(0,1,0)x(1,1,0)	Yes	0.1324
Yes	(0,1,0)x(0,1,1)	Yes	0.0525
Yes	(1,1,0)x(0,1,1)	Yes	0.2819
Yes	(0,1,0)x(2,1,0)	Yes	0.1059

Final Forecast with Errors:

Log	SARIMA Model (p,d,q)x(P,D,Q)	Sig. Param	Box Test	MAPE (n=1)	MAPE (n=2)	MAPE (n=3)	MAPE (n=4)
No	(0,1,0)x(1,1,0)	Yes	0.6719	6.370	8.612	8.985	9.022
No	(1,1,0)x(0,1,1)	Yes	0.2497	6.532	8.332	8.658	8.043
Yes	(0,1,0)x(1,1,0)	Yes	0.1324	5.280	7.382	9.064	9.624
Yes	(0,1,0)x(0,1,1)	Yes	0.0525	5.073	7.279	8.170	7.708
Yes	(1,1,0)x(0,1,1)	Yes	0.2819	5.691	7.116	7.541	7.123
Yes	(0,1,0)x(2,1,0)	Yes	0.1059	5.242	7.288	8.312	8.549

We chose the Rolling Scheme to assess the error, to avoid the potential problem with the model's stability, and it is more robust against structural breaks in the data as observed in our case when plotting the Coca-Cola earnings data.

The best model we chose for forecasting the Coca-Cola earnings time series is (0,1,0)x(2,1,0) because it has the second lowest MAPE (5.242) for n = 1. We did not choose the model (0,1,0)x(0,1,1) even though it has a lower MAPE (5.073) because we want a more robust model with a larger confidence when it comes to the residuals being white noise.

Below is a plot of our final model.