## Q1)

Seasonality is a time collection function wherein statistics report predictable versions that repeat each calendar year. Seasonal styles are described as predictable fluctuations that recur over one year. The results of seasonal companies encompass surprising seasonal versions in climate styles. An unexpected climate occasion may be catastrophic to even the best-laid commercial enterprise plans. We need to grasp and identify seasonal impacts, trends, and periods. It's the predictable swings in a one-year cycle. Cyclic patterns reflect data over a long term, usually collected yearly rather than for a set period. Cyclical patterns are usually affected by changes in economic conditions. A regression pattern occurs when the figures rise or decline over the longer term (*How seasonal forecasts can influence your business*). Figuring out seasonal variations, trends, and cycles, we identify the graph's seasonal variation.

Below the graph shows the seasonal trend as it has the same pattern observed every year.
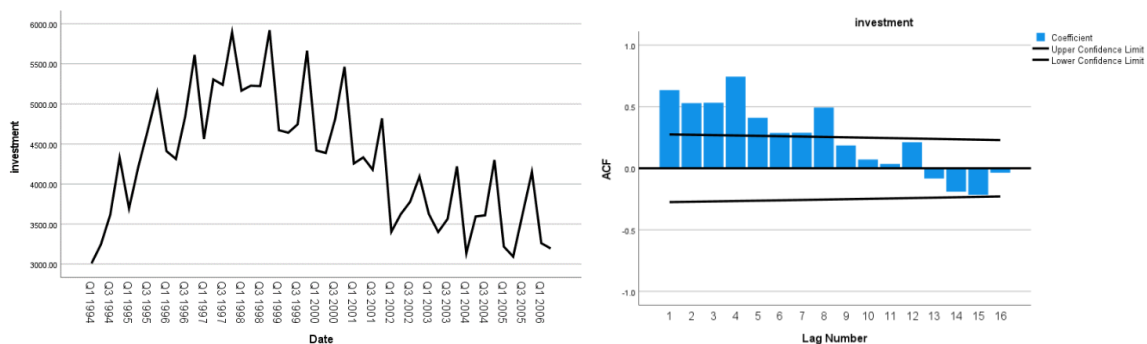
From the above graph we can see a seasonal effect sow as the pattern is repeating and we can see that in the graph from the year 1994 to 2002 there is an upward growth or trend but gradually from the year 2002 there is a downward trend for the next years. With this information, we can conclude that the following graph follows a cyclic pattern and seasonal pattern where there is a gradual increase and downfall at different stages. According to the above graph, we can notice there is a rapid increase in period four but gradually there is a decline of growth. From period 12, the growth of the industry has fallen to a negative limit.
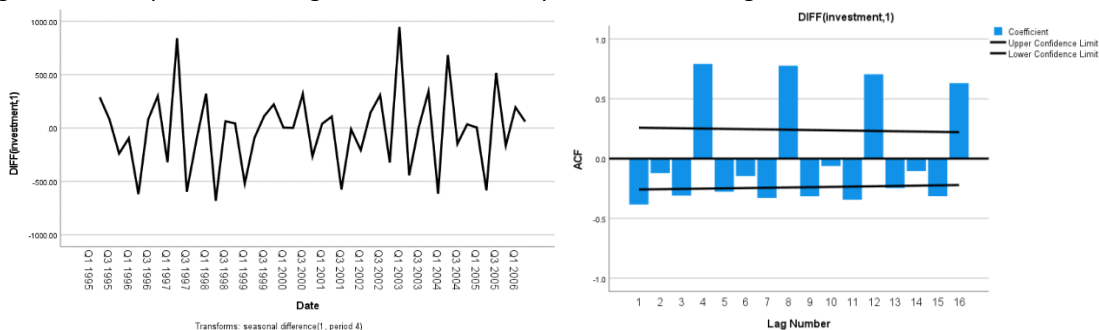
But to remove the trend line and make the seasonal effect clear, we take the difference out of raw data, which can show a graph like below, which has removed the trend cycle and shows the seasonal effect more clearly. From the ACF graph of the first difference, we can see that there is a spike and seasonality every fourth quarter.

We can conclude that it has strong seasonal effects by looking at the graph below taken from the first difference.

Q2) (a)

Business investment projection factors are regressed on time, Q1, Q2, and Q3. The model is not a good match, as seen in the SPSS result below. The adjusted R square value is 0.392, which indicates a poor fit. The regression coefficient has a significance of less than 5%, making it a good indication of the model. As a result, no regression analysis variable must be eliminated. 0.172 is the Durbin-Watson statistic. This illustrates the residual values' positive autocorrelation. The Durbin-Watson statistic with a value of 1.5 to 2.5 indicates no autocorrelation with successive data.

| Model Summary |
| --- |

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.665[a]	.442	.392	624.26969	.172
a. Predictors: (Constant), Q3, time, Q2, Q1					
b. Dependent Variable: investment					

Coefficients[a]

| Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5577.542	240.432		23.198	.000
	time	-23.466	6.121	-.427	-3.833	.000
	Q1	-1080.037	249.983	-.598	-4.320	.000
	Q2	-922.571	249.908	-.510	-3.692	.001
	Q3	-665.216	254.931	-.358	-2.609	.012
a. Dependent Variable: investment						

Below is the built dummy variable model

Business Investment=5577.542-23.466*time-1080.037*Q1-922.571*Q2-665.216*Q3 +error

The graph of modeling raw data Vs the corresponding forecast is shown below. The lines do not meet, and this also shows that the forecasted values have errors.
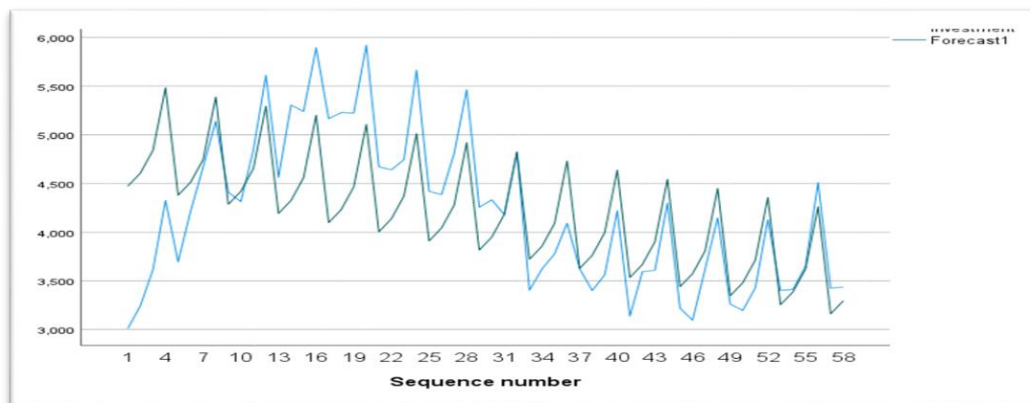
Sequence chart and ACF of the error are displayed below-
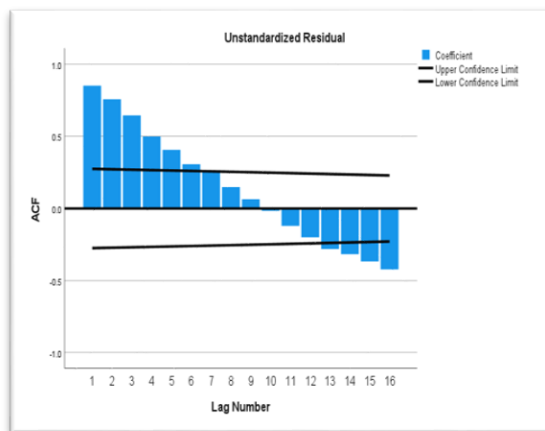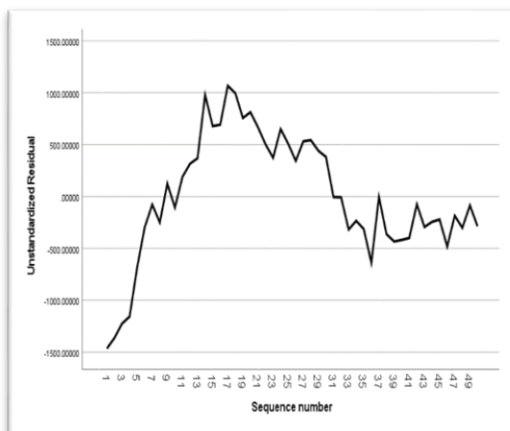
We can observe that errors are both positively and negatively balanced, with errors being larger than raw data. The inaccuracy is greater than 10% of the investment amount, which is unacceptable. In addition, the ACF displays many significant autocorrelations.

Q2) (b)

1. **Quadratic Trend Model -**

The value of time*time is produced to analyze the dummy variables method with the nonlinear trend-cycle component. The data is regressed on dummies time, time square(time*time), Q1, Q2, and Q3 to forecast the business investment. The result of SPSS is shown below.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.861[a]	.741	.711	430.22987	.328
a. Predictors: (Constant), Q3, time, Q2, Q1, time2					
b. Dependent Variable: investment					

| Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	4503.761	224.004		20.106	.000
	time	95.756	17.260	1.743	5.548	.000
	time2	-2.338	.328	-2.238	-7.124	.000
	Q1	-1002.114	172.628	-.554	-5.805	.000
	Q2	-844.647	172.577	-.467	-4.894	.000
	Q3	-665.216	175.691	-.358	-3.786	.000
a. Dependent Variable: investment						

The F-test p-value is 0.000, and the adjusted R-square is 71.1 per cent (still not a good match). As a result, no variable from the model will be eliminated. The Durbin-Watson value is less than 0.328. The Durbin-Watson Statistic score below 1.5 indicates a favorable correlation with subsequent data. The built dummy model is.
Business Investment= 4503.761+95.756*time-2.338*time2-1002.114*Q1-844.647*Q2- 665.216*Q3+ error
Following is the graph of modeling raw data Vs the corresponding forecast.
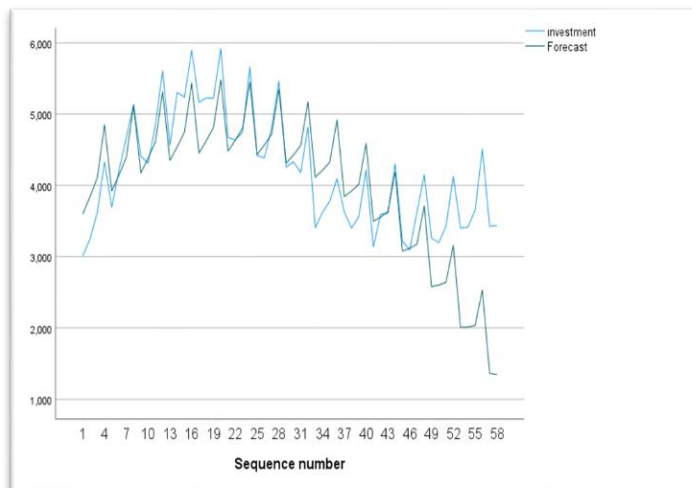
The raw data and forecast lines do not overlap, and they indicate some minor errors compared to the liner trend-cycle component's forecast. This model, however, does not satisfy us. As a result, we use the cubic trend value value to see if the graphs coincide. The sequence chart and the ACF are used to analyze these errors further.
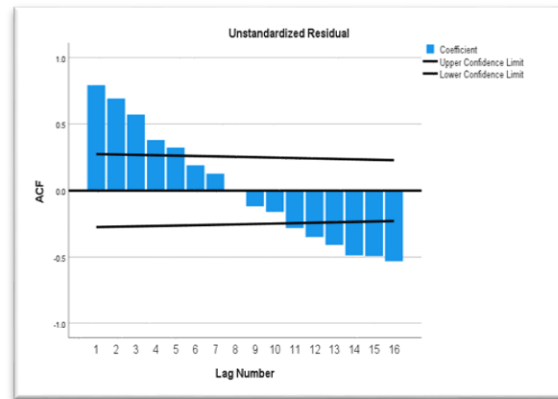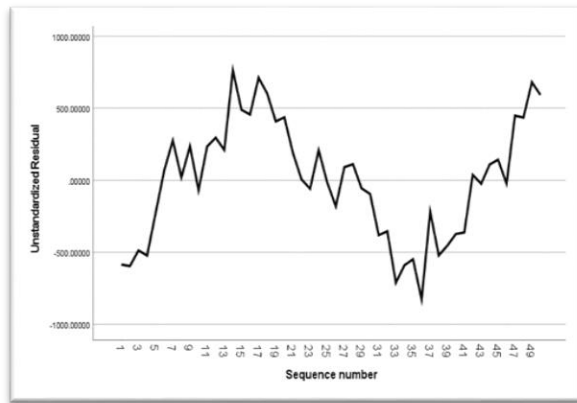
2. Cubic Trend Model-

The value of time*time*time is produced to analyze the dummy variables method with the nonlinear trend-cycle component as we were not satisfied with the square model. The data is regressed on dummies time, time square(time*time), time Cube(time*time*time), Q1, Q2, and Q3 to forecast the business investment. The result of SPSS is shown below.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.973[a]	.947	.940	196.40471	1.496
a. Predictors: (Constant), Q3, time, Q2, Q1, time3, time2					
b. Dependent Variable: investment					

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	3411.392	132.493		25.748	.000
	time	337.123	20.214	6.136	16.678	.000
	time2	-14.051	.916	-13.453	-15.345	.000
	time3	.153	.012	7.054	12.966	.000
	Q1	-964.157	78.861	-.533	-12.226	.000
	Q2	-852.303	78.785	-.472	-10.818	.000
	Q3	-634.915	80.239	-.342	-7.913	.000
a. Dependent Variable: investment						

The adjusted R-square is 94 percent (excellent fit), and the F-test p-value is 0.000. All regression coefficients are significant, and their value of less than 5% is regarded as a favorable indicator. As a result, no variable from the model will be eliminated. The Durbin-Watson value is less than 1.5, at 1.496. The Durbin-Watson Statistic score below 1.5 indicates a positive correlation with successive data.

The constructed dummy model is Business Investment= 3411.392-964.157*Q1-852.303*Q2-634.915*Q3+337.123*Time-14.051*TimeSquare+0.153*TimeCubic

Following is the graph of modeling raw data Vs the corresponding forecast.
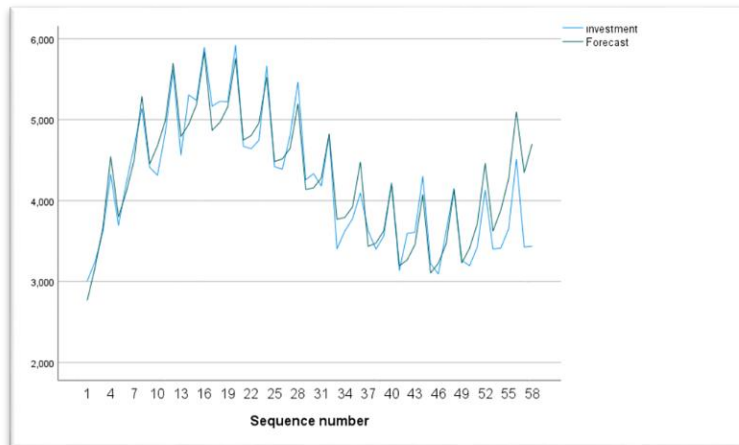
The raw data and forecast lines overlap, indicating some minimal errors compared to the liner trend-cycle and non-linear quadratic trend model component's forecast. This model satisfies the needs. As a result, we use the cubic trend model value because the graphs coincide. The sequence chart and the ACF are used to analyze these errors further.
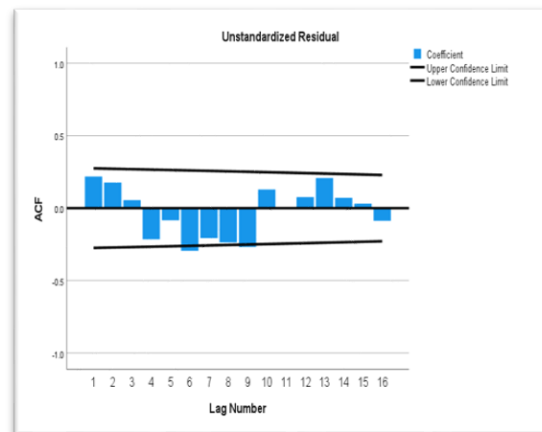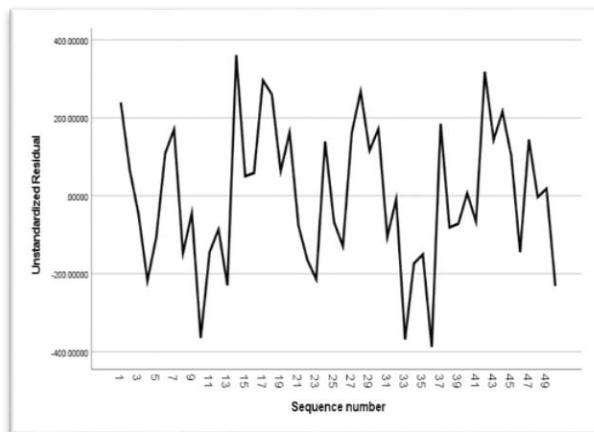

Q3)

Using the decomposition method, we will analyze the seasonal component of the data; we can find the trend cycle and irregular components by using the ARIMA model. We choose the additive rather than the multiplicative model for forecasting the values, as the analysis shows the seasonality component repeating every year.

Using the sequence chart of investment, we will check for the seasonality from the graph above.
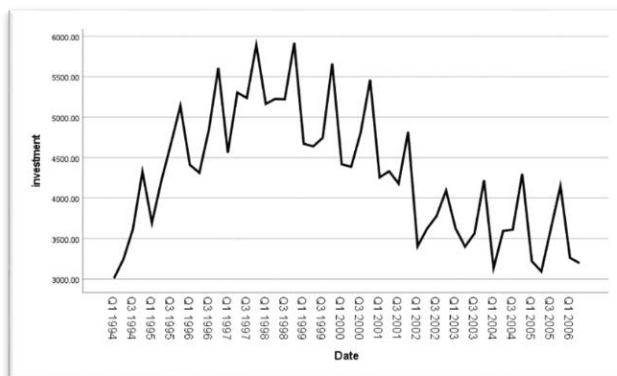
After careful examination of the pattern, the size of the seasonal and irregular components is reasonably constant throughout the data.

We can see a seasonal component from the above graph. By observing the irregular components throughout the data is constant, we choose an additive model for further forecasting of values. I will be using the Additive Model to forecast the values. We can show the additive model using graphs' seasonal, trend, and irregular components.

 Additive Model= Trend + Seasonal +Irregular.

To forecast the value, we will be needed to calculate seasonal factors and point forecast values. Therefore, this will involve three steps: -

STEP1- CALCULATION OF SEASONAL FACTOR

We have received the seasonal factors sequence chart and seasonally adjusted graph by using the decomposition method. According to the seasonally adjusted series graph, we can see that the values are not standardized. We are getting the first difference from the seasonally adjusted series to stabilize it. We can see that the highest seasonal value is the positive seasonal index value by using the seasonal index values. The deseasonalized data and seasonal index are displayed below-
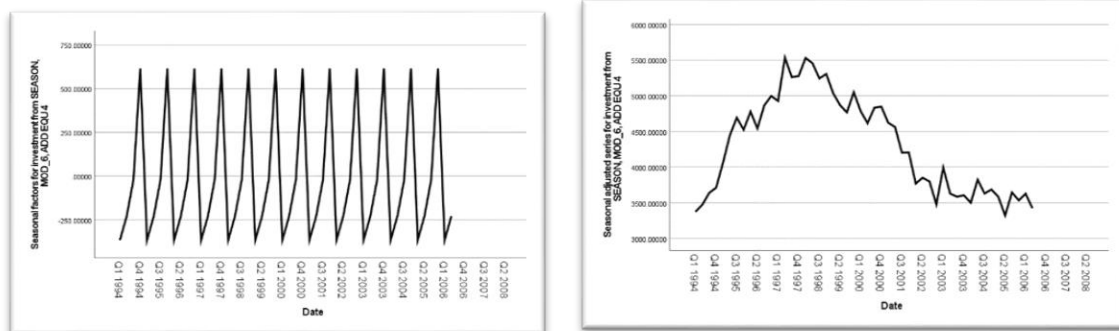
STEP2- Calculating the forecasting values

When calculating forecasting values, the initial and essential step to take while making the ARIMA model is to check whether the time series of the data is stationary. The data is stationary when the mean, variance, and autocorrelation structure of a stationary process do not change over time. We can predict stationarity through the run sequence in the graph by observing the deseasonalized time series, and we can predict and observe the strong trend cycle. Stationarity can represent a predictable pattern in the long term, constant variance over time, a constant autocorrelation structure, and no periodic fluctuations maintaining seasonality is called stationary (6.4.4.2 stationarity). The SAS's ACF and PACF graph is displayed below to further analyze the SAS time-series graph's stationery.
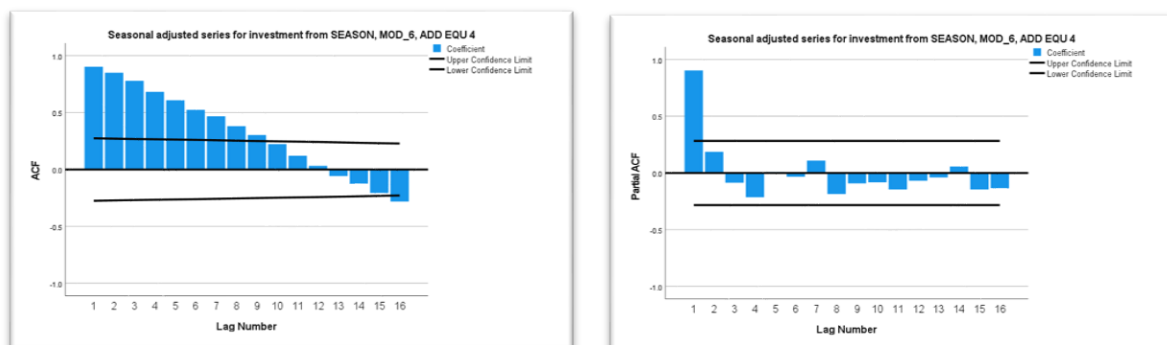
From the above graph, we can observe that the ACF graph shows a linear line as the values are gradually falling, and they are decaf or damped sinewave. When it comes to PACF, we can see a high increase of value in the Lag1 only. Therefore, we can conclude that the data of SAS raw data is not stationary and need to make a difference out of SAS values to get stationary values. To make the values stationary, we will do the first difference of SAS. The two graphs
below represent the time-series graph of SAS before difference and after the difference.
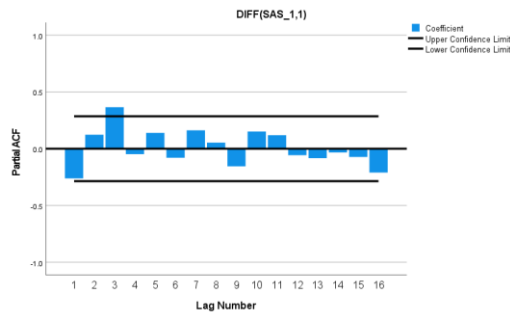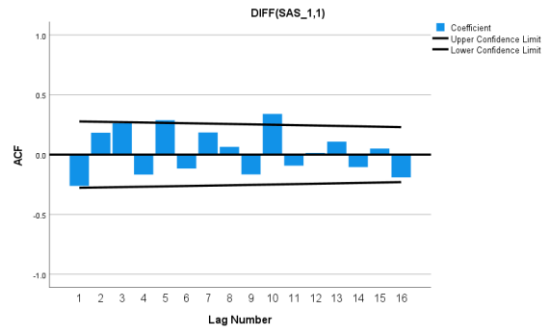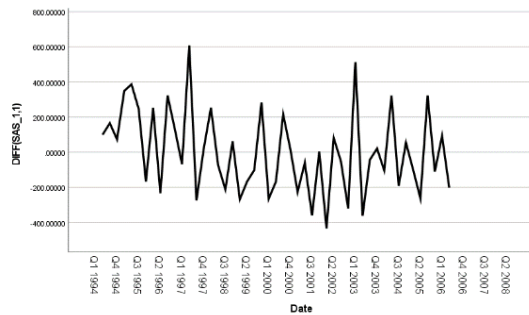
From the above graph, we can see after that the first difference has lost its linear trend. For further analysis, the ACF and PACF graph of the first difference SAS is displayed.

STEP 3: ARIMA MODEL

From the above graphs, we can conclude that both ACF AND PACF have shown damped sinewave and do not represent any pattern and insignificant autocorrelation which states it has stationarity in the time series and data. We got our correct time series by doing the first difference and can try predicting the models and analyzing them by checking both the AR and MR models (*Damped sine wave).*

STEP 3 ARIMA MODEL
From the graphs, we can see the possibilities for a plausible model such as ARIMA (0,1,5), (3,1,0), (1,0,0),(0,1,10) but we should select models on the criteria of their significant lag value from both ACF and PACF graphs but when it comes to (0,1,10) we have to use lag10 which is longer process might get things complicated while creating ARIMA model and assumed as an error and decided to stick to (0,1,5),(1,0,0) and (3,1,0). Hence, we will create ARIMA (1,0,0), ARIMA (3,1,0), ARIMA (0,1,5). These models are compared based on their component p-value, Normalized BIC. Through overfitting, by increasing the p and q by 1 we can pick the best models.

Arima (1,0,0)

Model Statistics								
		Model Fit statistics			Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	0	.868	.868	11.224	23.890	17	.122	0

From the above table, we can see the normalized BIC is 11.224 and from the Ljung box the significane and pvalue is 0.122 which is not significant.

ARIMA Model Parameters	Estimate	SE	t	Sig.

Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4	No Transformation	Constant		3937.066	487.461	8.077	.000
			AR	Lag 1	.945	.043	22.016	.000

The AR lag1 value is significant with the value 0.

Arima (3,1,0)

Model Statistics

Model	Number of Predictors	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	0	.208	.900	11.132	12.629	15	.631	0

From the above graph, we can see the normalized Bic is 11.132 which is lesser than the ARIMA model (1,0,0) and from ljung box statistics thep-valuee is 0.631 which is also not significant.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4	No Transformation	Constant		3.660	43.678	.084	.934
			AR	Lag 1	-.270	.138	-1.951	.057
				Lag 2	.208	.141	1.477	.147
				Lag 3	.357	.139	2.562	.014
			Difference		1			

The AR Lag 3 value is significant with a p-value of 1.4%. the estimated coefficient withp-valuee is less than 5%.

Arima (0,1,5)

Model Statistics

Model	Number of Predictors	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	0	.231	.903	11.307	12.507	13	.487	0

From the model statistics we can see that the normalized BIC is 11.307 and from ljung box we can see the p-value is 0.487 which is not significant.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4	No Transformation	Constant		2.448	40.511	.060	.952
			Difference		1			
			MA	Lag 1	.277	.148	1.872	.068
				Lag 2	-.292	.152	-1.925	.061
				Lag 3	-.267	.153	-1.745	.088
				Lag 4	.303	.151	2.005	.051

				Lag 5	-.323	.151	-2.133	.039

From the above details, we can see the MA Lag 3 has a significant p-value by 3.9% and which is less than 5%. After comparing the Normalized Bic from above all models we can predict that ARIMA (3,1,0) has the least BIC and the AR Lag 3 is also significant in ARIMA (3,1,0).

Now we will do the overfitting by estimating the ARIMA model with higher-order, we are estimating on ARIMA(3,1,0) model. By

a) Increasing the order p by 1- ARIMA (4,1,0)
b) Increasing the order q by 1- ARIMA (3,1,1)

ARIMA (4,1,0)

Model Statistics								
		Model Fit statistics			Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	0	.209	.900	11.233	12.819	14	.541	0

The Normalized BIC is 11.233 which is higher than the BIC of the AR3 in the first difference. The statistic of the Ljung Box Hypothesis is also not significant.

ARIMA MODEL PARAMETERS					Estimate	SE	t	Sig.
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4	No Transformation	Constant		2.845	42.855	.066	.947
			AR	Lag 1	-.256	.153	-1.676	.101
				Lag 2	.214	.146	1.465	.150
				Lag 3	.347	.147	2.366	.022
				Lag 4	-.034	.156	-.220	.827
			Difference		1			

In the above table, the AR Lag4 is not significant, and the p-value is more than 5% significance.

From this, we can conclude that ARIMA (3,1,0) is the best model if we remove the Lag4 from the above model ARIMA (4,1,0).

ARMA (3,1,1)

Model Statistics								
		Model Fit statistics			Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	0	.216	.901	11.223	14.280	14	.429	0

From the above table, we can see the Bic is 11.223 which is higher than the Bic of ARIMA (3,1,0). The ljung box is not showing a significant value and shows a 0.42p-valuesue.

ARIMA Model Parameters					Estimate	SE	t	Sig.
Seasonal adjusted series for investment	Seasonal adjusted series for investment	No Transformation	Constant		1.132	38.930	.029	.977
			AR	Lag 1	-.708	.297	-2.381	.022

from SEASON, MOD_2, ADD EQU 4-Model_1	from SEASON, MOD_2, ADD EQU 4			Lag 2	.096	.190	.507	.615
				Lag 3	.382	.143	2.676	.010
			Difference	1				
			MA	Lag 1	-.523	.318	-1.645	.107

We can see the MA value of Lag1 is not significant and the p-value is above the significance level of 5%. If we remove the MA Lag1 from ARIMA (3,1,1) we can get the selected module (3,1,0) which is the best and most suitable model.

After comparing the overfitting models and the plausible model by comparing the Bic and other factors we can see that the other models as in ARIMA (3,1,1) AND ARIMA (4,1,0) we can see that these models have higher Normalized Bic compared to ARIMA (3,1,0). Through this, we can conclude that ARIMA (3,1,0) is the best model suitable for our problem.

ARIMA (3,1,0)

| Model Statistics								
		Model Fit statistics			Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
Seasonal adjusted series for investment from SEASON, MOD_2, ADD EQU 4-Model_1	0	.208	.900	11.132	12.629	15	.631	0

From the above graph ARIMA (3,1,0) we can see the r-square value .900 where the variability of deseasonalized data is 90% which is good, and the stationary r square is .208 which is 20.8% which is also a good value after the first difference ARIMA(3,1,0).

By using formula $U_t = y_t - y_{t-1}$ from AR3 of first difference we can drive a forecasting equation to the build model to deseasonalized investment:

$u_t - 3.660 = (-0.270) *(u_{t-1} - 3.660) + 0.208(u_{t-2} - 3.660) + 0.357(u_{t-3} - 3.660) + e_t$

The forecasted value of the ARIMA (3,1,0) is compared with the actual deseasonalized data.

ERROR ANALYSIS: We check residual errors by using the formula residuals errors = actual value- forecast value we can check if the pattern remains or not by using this.
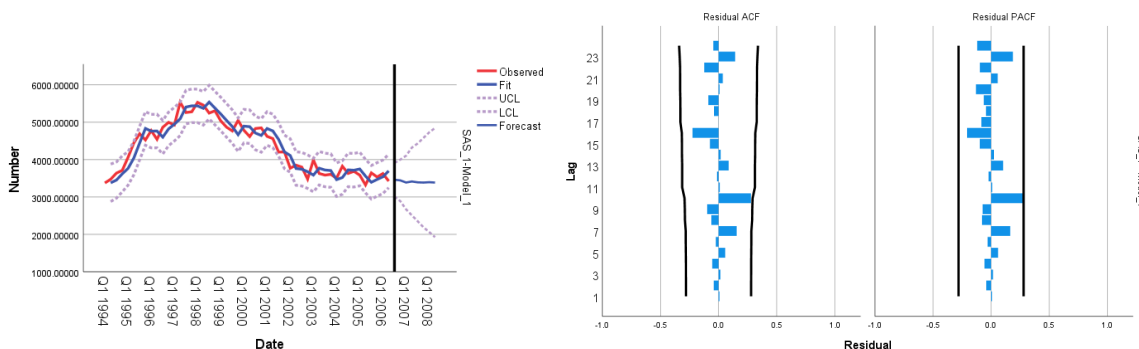
Through the process above, we can see that the forecasted data has some residual errors, and this residual has been analysed by using the ACF and PACF graph. We can see no significant pattern and the values of ACF and
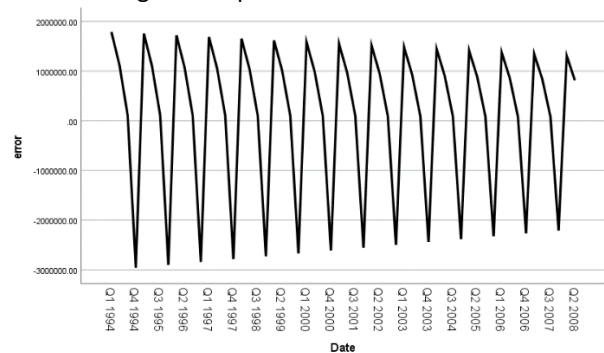
PACF are also not significant.

we can observe from the graph representing white noise as well. From the graph, we can present the error analysis data obtained which shows a similar pattern throughout the year but with slight improvement can be seen. Whereas error for investment has a similar pattern with ups and downs. For a good forecasting model, the residuals leftover after overfitting should contain only white noise. We can say residual values have been balanced with positive and negative values randomly. Hence, we analyzed the errors through time series and ACF and PACF graphs. We can find no significant autocorrelation and we couldn't find the significance in error

Predicted SAS	SAF	FORECAST VALUES	HOLDBACK DATA	error
3461.59	-21.34612	3440.24388	3430	10.24388
3444.97423	615.15388	4060.12811	4130	-69.87189
3387.28939	-365.30445	3021.98494	3403	-381.01506
3415.2355	-228.50331	3186.73219	3414	-227.26781
3392.31758	-21.34612	3370.97146	3652	-281.02854
3386.2872	615.15388	4001.44108	4511	-509.55892
3395.7081	-365.30445	3030.40365	3426	-395.59635
3386.29686	-228.50331	3157.79355	3437	-279.20645

analysis.

Forecasting values of holdback data:

Q4)

Three models are compared: a dummy variable model with a linear trend-cycle component, a dummy variable model with a non-linear trend-cycle (Cubic) component, and a combination of the decomposition method and the Box-Jenkins ARIMA approach. We'll look at the modelling data and the hold-back data prediction to see if the forecast is accurate. The R-square of differences will be compared during the analysis.

Model Name	R-squared	Adjusted R-square	Stationary R-square
Dummy variable model with linear trend-cycle component	0.442	0.392	
Dummy variable model with the non-linear trend-cycle component (Cubic)	0.947	0.940	
Box-Jenkins ARIMA (3,1,0)	0.900		0.208

We'll use adjusted r-square instead of R-square to assess the dummy variable models since R-square overstates the theoretical value of the coefficient of determination. Adjusted r-square, on the other hand, minimizes the estimator R-square's bias. In addition, adding an independent variable to the model invariably raises the r-square value every time (Frost, 2017). As a result, the adjusted R-square value should be used to compare the dummy model. The non-linear trend-cycle component's adjusted R square is comparatively high, indicating a superior model. Based on R-square, the Dummy Variable model with the non-linear trend-

cycle component is compared to the Box-Jenkins ARIMA (3,1,0). The R-square value is compared because it measures the error and helps us anticipate the value of the dependent variable.

More errors are represented by a smaller R-square value (Kpolovie, 2017). With careful examination, Box Jenkins ARIMA (3,1,0) has more minor errors than R Square's 90% value. When all three models in the sample forecast were compared, Box Jenkins ARIMA (3,1,0) showed to be the best. The MAD (Mean Absolute Deviation), MSE (Mean Squared Error), and MAPE (Mean Absolute Percentage Error) values will be used to analyze the holdback data. Checking the out-of-sample forecast is a better guide to evaluating the different models. This evaluation of out-of-sample data will assist us to grasp the importance of our model in predicting future values.

Dummy Variable Model with Linear Trend

Holdback data from SPSS	Forecasts from SPSS	Forecast errors	Squared errors	Abs errors	Abs relative errors
		formula=B14-C14	formula=D14*D14	formula=abs(D14)	formula=100*abs(D14)/B14
3430	3715.56	-285.56	81544.5136	285.56	8.3254
4130	4357.31	-227.31	51669.8361	227.31	5.5039
3403	3253.81	149.19	22257.6561	149.19	4.3841
3414	3387.81	26.19	685.9161	26.19	0.7671
3652	3621.7	30.3	918.09	30.3	0.8297
4511	4263.45	247.55	61281.0025	247.55	5.4877
3426	3159.94	266.06	70787.9236	266.06	7.7659
3437	3293.94	143.06	20466.1636	143.06	4.1624
Calculate Mean			38701.3877 **MSE**	171.9025 **MAD**	4.6533 **MAPE (%)**

Dummy Variable Model with Non-Linear Trend

Holdback data from SPSS	Forecasts from SPSS	Forecast errors	Squared errors	Abs errors	Abs relative errors
		formula=B14-C14	formula=D14*D14	formula=abs(D14)	formula=100*abs(D14)/B14
3430	3718.7	-288.7	83347.69	288.7	8.4169
4130	4460.91	-330.91	109501.4281	330.91	8.0123
3403	3623.68	-220.68	48699.6624	220.68	6.4849
3414	3883.01	-469.01	219970.3801	469.01	13.7378
3652	4269.34	-617.34	381108.6756	617.34	16.9042
4511	5095.59	-584.59	341745.4681	584.59	12.9592
3426	4346.08	-920.08	846547.2064	920.08	26.8558
3437	4696.79	-1259.79	1587070.844	1259.79	36.6538
Calculate Mean			452248.9194 **MSE**	586.3875 **MAD**	16.2531 **MAPE (%)**

Decomposed ARIMA (3,1,0)

Holdback data from SPSS	Error of Decomposed ARIMA model	Squared errors	Abs errors	Abs relative errors
3430	-10.25	105.063	10.25	0.30
4130	69.87	4881.81	69.87	1.70
3403	381.02	145176.20	381.02	11.20
3414	227.27	51651.65	227.27	6.65
3652	281.03	78977.86	281.03	7.69
4511	509.56	259651.40	509.56	11.29
3426	395.6	156499.40	395.6	11.55
3437	279.21	77958.22	279.21	8.12
Calculate Mean		96862.7 **MSE**	269.2263 **MAD**	7.31 **MAPE (%)**

When the three models are compared, the Dummy Variable Model with Linear trend has the best fit, with the lowest MAD, MSE, and MAPE values. Because MAD stands for the absolute average of error, which considers whether the value was overestimated or underestimated. The smaller MSE indicates the model's stability, but it may also be deceptive because it emphasizes a significant error term. Because MAPE does not emphasize major error terms, it is the optimum error quantification (Weatherford, et al., 2003).

Though the values are more minor, the linear model is not well-suited if we consider in-sample validation. A substantial linear-trend pattern was detected based on in-sample error analysis, as we mentioned in question 2. According to the Durbin Watson Statistic, it also exhibits a positive autocorrelation in the residual value. The model may have to rely on chance to predict the holdback, but it does not mean it is effective (Duke FUQUA, 2021). As a result, the dummy variable model with the linear trend cannot be used to anticipate future values since in-sample validation does not provide an acceptable match.

The combined technique of the decomposition model and the Box-Jenkins ARIMA Approach is the 2nd best model for comparing out-of-sample forecast error. The MAPE result of 7.31 percent indicates that residual error was minimal and controllable. The combination technique produced improved findings after combining the study of in-sample validation. The model is quite successful and has proved its efficacy throughout the modelling sample, according to error analysis. The residual ACF graph indicates no notable patterns. The model's R-square was also the greatest, with a 90 percent variability to the decomposed data. We shall disregard the value of the stationary R-square because the data after de-personalization shows a clear trend. Because it measures the variability of the data after eliminating the trend-cycle component, R-square is a suitable method of comparing different ARIMA models. Hence, ARIMA (3,1,0) combined with the decomposition model is the best method to forecast future values.

Q5)

Autoregressive integrated moving average (ARIMA) models forecast future values based on historical data. ARIMA smooths time series data using lagged moving averages. Autoregressive models believe that the future will be the same as the past. As a result, in some market situations, they may be inaccurate. We decided to use a combination method of decomposition model and box Jenkins model to select the ARIMA model as the most suitable method to find the ARIMA model after a lot of studies and practice because this method will help to estimate the seasonal effect and the ARIMA model will help to estimate the linear trend and irregular components. The final chosen model is the combination method of the decomposition model and the Box Jenkins ARIMA (3,1,0) Approach (Hayes, A. (2022)).

One of the primary goals of decomposition is to estimate seasonal effects that can be used to generate and display seasonally adjusted values. A seasonally adjusted value removes the seasonal effect from a value, making trends more visible. After careful examination, the size of seasonal and irregular components was assumed to be constant throughout the data. This assumption was critical in choosing the model to forecast the investment amount (*5.1 decomposition models*).

We have two types of models called the additive model and multiplicative model. Choosing a model directs us to the formula for calculating the forecast value and explaining the seasonal index's nature. The sum of the seasonal index in the multiplicative model is 4, whereas the sum of the seasonal index in the additive model is 0, so we chose the additive model. We can see constant seasonal and irregular components in the data.

As the seasonal pattern is repetitive, we can collect the data at constant time intervals, which helps us examine the seasonal pattern required. As per the graphs, we can see there are good spikes which tell there was a boom period. We can assume the dependent and independent variables time series, interval time where time point is constant throughout the data.

The difference value showed the trend component, and the ARIMA approach may estimate the irregular components using the stationary model. The linear trend is represented by first-order differencing, whereas the quadratic trend is represented by second-order differencing. The data show a clear downward linear trend with the study of seasonal, trend, and cycle impacts in the first portion.

After the first difference, the deseasonalized time series has become stationary for the custom ARIMA model. The fixed time series can maintain its statistical average and self-degradation structure over time. After examining the raw data from the seasonally adjusted series, we couldn't find apparent seasonal effects, so we took the first difference, removing the trend and giving the seasonal data a clear and suitable format for predicting an ARIMA model.

From the first difference, we have taken a few models where we select the model when the data is stationary, and the normalized Bic is significantly less. We can overfit the model by adding AR and MA values and check the model if it's correct or not, as the AR and MA value specifies the dataset and forecast values for the future. If we take a second difference for making the ARIMA model still, the data can be more constant and seasonal, and the changes in AR and MA would be significant. So even if the assumptions are violated, we still get the correct model with the suitable AR and MA values if we select the right model.

Through ARIMA Model, AR indicates that the value of the previous three years is only relevant to forecasting the value, and MA indicates the previous three years' value won't be sufficient to create an error. Thus, ARIMA (3,1,0) is the best suitable model.

<div align="center">Reference List</div>

5.1 decomposition models (no date) *PennState: Statistics Online Courses*. Available at: https://online.stat.psu.edu/stat510/lesson/5/5.1 (Accessed: May 10, 2022).

6.4.4.2. Stationarity (no date) *Nist.gov*. Available at: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm (Accessed: May 9, 2022).

Damped sine wave: Definition, example, formula (2021) *Calculus How To*. Available at: https://www.calculushowto.com/calculus-definitions/damped-sine-wave/ (Accessed: May 10, 2022).

Fuqua School of Business (no date) *How to compare models*, *Duke.edu*. Available at: https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/compare.htm (Accessed: May 11, 2022).

Frost, J. (2017) *How to interpret adjusted R-squared and predicted R-squared in regression analysis*, *Statistics By Jim*. Available at: https://statisticsbyjim.com/regression/interpret-adjusted-r-squared-predicted-r-squared-regression (Accessed: May 11, 2022).

Hayes, A. (2022) *Autoregressive Integrated Moving Average (ARIMA)*, *Investopedia*. Available at: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp (Accessed: May 10, 2022).

How seasonal forecasts can influence your business (no date) *Altametrics.com*. Available at: https://altametrics.com/en/sales-forecast/seasonality-forecasting.html (Accessed: May 10, 2022).

Makridakis, S. G. and Wheelwright, S. C. (1983) *Forecasting: Methods and Applications*. 2nd ed. Brisbane, QLD, Australia: John Wiley and Sons (WIE).

Weatherford, L. R., Gentry, T. W. and Wilamowski, B. (2003) "Neural network forecasting for airlines: A comparative analysis," *Journal of Revenue and Pricing Management*, 1(4), pp. 319–331.

What is a Scatter Diagram? (no date) *Asq.org*. Available at: https://asq.org/quality-resources/scatter-diagram (Accessed: May 10, 2022).