Forecasting The Sales Volume Of Seven-Up Bottling Company Plc, Kaduna Plant Using The Classical Decomposition Method

Uwa, Celestine Ozoemenam[1] and Ogwueleka, Francisca Nonyelum[2]

Computer Science Department, Nigerian Defence Academy, Kaduna[1,2]

Abstract: The Seven-Up Bottling Company Plc, Kaduna Plant has no mathematical model that takes into account seasonal indexes for forecasting sales volume and since their products are seasonal in nature, the classical decomposition of time series data into the various components was used to model the present using the past trends into the future periods in order to have good forecasts for effective management. Sales records of Seven-Up Bottling Company Plc, Kaduna Plant for the past ten years (2007 – 2016) were collected, analysed and then various statistics such as t-test, z-test and Durbin-Watson test were carried out for the purpose of testing hypothesis. A model was then developed using deseasonalised data series that accounted for 95% of the total variability. The results showed that the seasonal indexes for the months of January, February, March, April, May and December were higher than those of the months of June, July, August, September, October and November. The highest seasonal value of 113.8% was recorded in May and the lowest of 75.3% was recorded in August. The model was used to forecast product sales for 2017 with the result showing good agreement between predicted and observed values.

Keywords: Classical decomposition; autocorrelation coefficient; seasonal indexes; trend-cycle; irregular component; cyclical component; trend seasonal factors.

I. INTRODUCTION

Sales forecasting is the process of estimating future sales (Susan, 2017). Accurate sales forecasts enable companies to make informed business decisions and predict short-term and long-term performance. Companies can base their forecasts on past sales data, industry-wide comparisons, and economic trends. It is easier for established companies to predict future sales based on years of past business data. Newly founded companies have to base their forecasts on less-verified information, such as market research and competitive intelligence to forecast their future business. Sales forecasting gives insight into how a company should manage its workforce, cash flow, and resources. In addition to helping a company allocate its internal resources effectively, predictive sales data is important for businesses when looking to acquire investment capital (Susan, 2017).

One of the oldest commonly used forecasting methods, the Classical Time Series Decomposition, will be used here because it includes the concept of seasonal indexes. This concept is important because Seven-Up products just like many other business and economic time series are seasonal and seasonal indexes are essential part of effective forecasting methods.

II. REVIEW OF RELATED LITERATURE

2.1 Sales Forecasting

A sales forecast predicts what a salesperson, team, or company will sell in a given time period - weekly, monthly, quarterly, or annually. Managers use their representatives' individual sales forecasts to estimate how much business their entire team will close. Directors use team forecasts to anticipate sales for the entire department. The Vice President of Sales uses department forecasts to project sales for the entire organization. These reports are typically shared with company leadership, along with board members and/or stockholders (Aja, 2017). The process of forecasting is based on reliable data of past and present. Forecasting is not new, as it has been practiced from time immemorial (Shreyasi, 2016). Accurate sales forecasts enable companies to make informed business decisions and predict short-term and long-term performance. Companies can base their forecasts on past sales data, industry-wide comparisons, and economic trends. It is easier for established companies to predict future sales based on years of past business data. Newly founded companies have to base

their forecasts on less-verified information, such as market research and competitive intelligence to forecast their future business (TrackMaven, 2017).

2.2   Classes of Sales Forecasts

There are three classes of sales forecasts which modern organizations require depending on the specific application: short-term, medium-term and long-term forecasts.

i.   **Short-term forecasts:** These are needed for the scheduling of personnel, production and transportation (Rob and George, 2014). It is also known as operating forecast, covering a maximum of one year or it may be half-yearly, quarterly, monthly and even weekly. Short-run forecast is preferred to all types and brings more benefits than other types (Shreyasi, 2016).
ii.   **Medium-term forecasts:** This type of forecast may cover from more than one year to two or four years. This helps the management to estimate probable profit and control over budgets, expenditure, production etc. (Shreyasi, 2016). These are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment (Rob and George, 2014).
iii.   **Long-term forecasts:** This type of forecast may cover one year to five years, depending on the nature of the firm. It is good for adding new products and dropping old ones (Shreyasi, 2016). These are used in strategic planning. Such decisions must take account of market opportunities, environmental factors and internal resources (Rob and George, 2014).

2.3   Forecasting Data and Methods

The appropriate forecasting methods depend largely on what data are available. The three forecasting methods based on the availability of data are: qualitative method, used if there are no data available, or if the data available are not relevant to the forecasts (Rob and George, 2014); quantitative method, applied when the numerical information about the past is available (Rob and George, 2014) and causal models which uses highly refined and specific information about relationships between system elements, and is powerful enough to take special events formally into account (John et al, 2007). The quantitative method is used here because of the availability of past numerical data and also the assumption that some aspects of the past patterns will continue into the future.

III.   DATA COLLECTION AND METHODOLOGIES

3.1   Data collection
The research data for this work was gathered through secondary and primary sources. The methods used in collecting data were document review, interview and observation. All the necessary materials were gathered from existing records, some key staff of the Company were interviewed and processes observed. All the manuals and documents relating to the daily, weekly, monthly, and annual sales figures were carefully examined and the required information extracted. Interview method was used to gather information from some key staff of the company. The information obtained through observation conforms significantly with what was obtained during the interview session.

3.2   Classical Decomposition

It is a relatively simple procedure and forms the basis for most other methods of time series decomposition. In classical decomposition, we assume the seasonal component is constant from year to year and are sometimes called the "seasonal indices" There are two forms of classical decomposition: an additive decomposition and a multiplicative decomposition. (Rob and George, 2014). In this work, the multiplicative model given by $Y = TCSe$ is used because it yields several simple identities for estimating the individual components. Here, T is the trend component, C, is the cyclical component, S, is the seasonal variation and the error (e) is the residual component (Rob and George, 2014).

IV.   ANALYSIS OF DATA

The Statistical Package for Social Sciences (SPSS) was used to analyze the data on Appendix A which shows the sales of SevenUP Bottling Company Plc Kaduna Zone from 2007 to 2016.

4.1   The 12 Months Moving Average

The calculation of the 12 months moving average was performed by the Software (SPSS). Table 1 shows the moving averages for the period, 2007 to 2016.

Table1: Sales volume of Seven-Up Kaduna Zone (2007 - 2016) 12 months moving average.

Month	2007	2008	2009	2010	2011
Jan	.	177368.6667	168253.3333	213023.4167	233656.0000
Feb	.	180015.8333	167906.0833	215635.4167	234380.1667
Mar	.	182546.7500	166290.6667	217700.2500	235596.0833
Apr	.	185223.3333	167581.5833	218843.4167	236830.0833
May	.	187687.3333	168872.4167	222723.3333	235270.0000
Jun	.	187262.0833	172000.1667	223674.4167	233481.0833
Jul	176332.5000	185984.5833	175268.4167	224132.5000	235929.0000
Aug	178013.5000	182924.9167	180673.7500	225901.4167	239158.4167
Sep	178296.5000	179624.8333	187034.5000	227758.8333	241571.9167
Oct	178964.2500	176099.3333	193473.7500	229513.6667	243901.2500
Nov	179674.1667	172944.3333	199587.7500	231239.5833	246603.0833
Dec	180243.7500	168719.3333	207081.4167	232907.1667	249140.7500
	2012	2013	2014	2015	2016
Jan	250557.5833	261046.8333	268388.3333	303103.2500	301881.8333
Feb	250942.1667	262243.5833	271049.0000	302226.7500	316949.2500
Mar	252342.1667	262468.8333	271897.6667	304053.7500	322429.0833
Apr	251442.5833	264192.6667	275686.4167	301072.5833	327972.4167
May	250736.5000	265947.0000	280369.6667	298455.3333	3331807500
Jun	251127.9167	267030.5833	285553.3333	299026.5833	336177.9167
Jul	253687.1667	267039.1667	287138.9167	302251.5000	335835.0000
Aug	253736.4167	268540.0000	289605.5000	302754.6667	.
Sep	254278.5833	269520.8333	292092.3333	303988.9167	.
Oct	255117.4167	270622.5833	294376.7500	305013.9167	.
Nov	256266.2500	271130.8333	296615.3333	306738.8333	,
Dec	257481.2500	271549.1667	299100.7500	308063.5000	.

4.2   Seasonal indexes (SI) % monthly

Table 2 shows the seasonal factors for the months of January to December as 110.8% to 108.0%. It should be noted that the weather is always very hot during these months with high seasonal index (SI) (%) in this part of the country.

Table 2: Shows the seasonal indexes (SI) % monthly

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
SI %	110.8	112.4	112.7	113.4	113.8	99.6	84.6	75.3	84.9	88.4	96.1	108.0

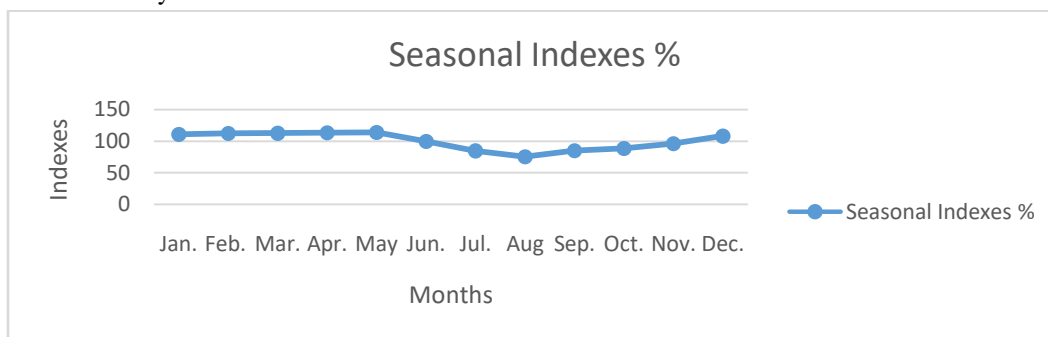Figure 1 shows the monthly seasonal indexes %.

Figure 1 Graph of Seasonal Indexes

4.3   Trend - cycle

Figure 2: Shows the graph of trend-cycle obtained by deseasonalizing the actual values (data).
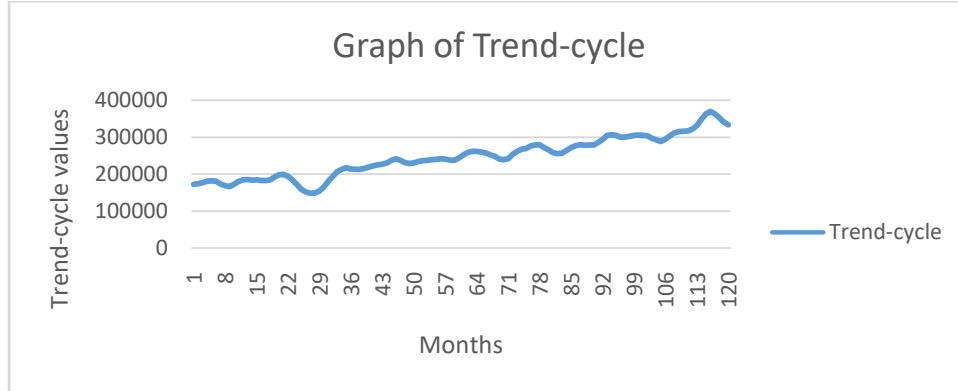

Figure 2: Graph of Trend-cycle

4.4   Cyclical Component

The lowest value for the cyclical factor occurred in May 2011 with C value of 81.45% while the highest cyclical value of 177.86% was recorded in September 2016.

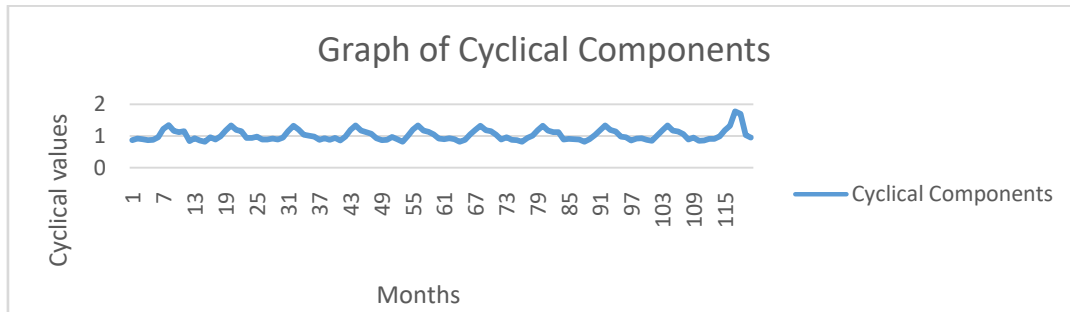Figure 3 shows the cyclical components.

Figure 3: Graph of Cyclical Components

4.5   Trend seasonal factors or fitted values ($\hat{\mathbf{Y}}$)

To obtain the trend seasonal value or fitted value $\hat{\mathbf{Y}}$ of the trend, T, we multiply it by its corresponding seasonal factor S_i. The result shows that the greatest fitted values occur with low values of S_i and lower values occur with high seasonal factors of S_i. This is shown in Figure 4.

Figure 4 shows the trend seasonal or fitted values of the data.
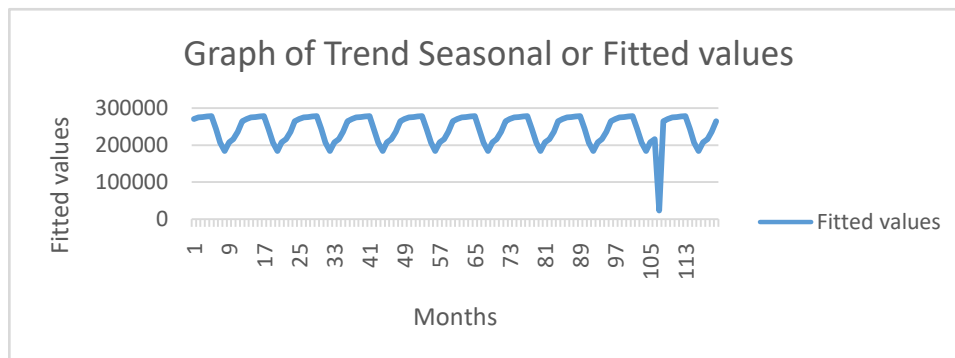

Figure 4: Graph of Trend seasonal or fitted values

Figure 5 shows the graph of fitted values (trend seasonal) against original values. The figure indicates the closeness of the two values and hence, the model if adopted by the company will serve the Company's forecasting needs.
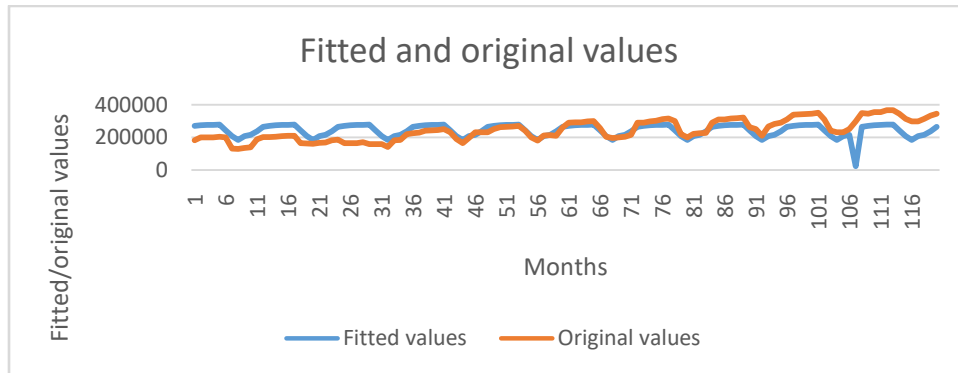

Figure 5: Graph of fitted (trend seasonal) and original values.

4.6   Research Hypothesis

4.6.1   Research Questions

The research questions are:

a.   What is the nature of the patterns in the data series concerning: trend, seasonal, cyclical   and irregular patterns?

b.   Are there any observed autocorrelation in this series? Is the autocorrelation positive or negative?

c.   How acceptable is the goodness of fit for the model?

d.   What will be the confidence interval of the estimation as a measure of forecast uncertainty using the regression line?

e.   Are the coefficients of the regression model statistically significant?

f.   Are there outliers in the data base that can cause bias?

g.   Are the white noise Autocorrelation Coefficient [ACF(k)], Partial Autocorrelation Coefficient [PACF(k)] statistically significant?

h.   How statistically significant are the Cross-correlation Coefficients (CCF)?

i.   In estimating the trend in the decomposition method using linear regression methods, where lies a more accurate result?

4.6.2   Research Hypothesis

To answer the questions raised in section 4.6.1, the following hypothesis will be tested during the analysis of the result.

a.   With regard to the question of whether there is a systematic bias in a forecast ($\bar{e} \neq 0$) to confirm that the mean error is statistically different than zero.

H_0: $\tilde{e} = 0$ means there is no bias

H_1: $\tilde{e} \neq 0$ means there is a bias in the trend.

b.   With respect to the significance of regression coefficient b, the following hypotheses is stated.

H_0: $\beta = 0$ means the true regression coefficient (β) is not statistically significantly different from zero.

H_1: $\beta \neq 0$ means the true regression coefficient (β) is statistically significantly different than zero.

c.   The Covariance will be used to determine the strength of the linear relationship between the variables x and y. The hypotheses to be tested are:

$H_0 : \rho_{XY} = 0$ means there is no linear relationship between the two variables.

$H_1 \rho_{XY} \neq 0$ means there is a linear relationship between the two variables.

d.   The Durbin-Watson statistic will be used to conduct a t-test for autocorrelation significance as follows:

H_0: $e_t = 0$ the errors e_t are auto correlated with first order scheme.

H_1: $e_t \neq 0$ the errors e_t are not auto correlated with a first order scheme.

e.   The CCF t-test will be used to test for statistical significance of cross-correlation as follows:

H_0: CCF = 0 means that the cross-correlation coefficient are not statistically significant.

H_1: CCF $\neq$ 0 means that the cross-correlation coefficient are statistically significant.

4.6.3   Statistical test for the bias for the model.
A simple t-test was carried out and t-calculated is -5.25. From standard t-table, degrees of freedom = n-2-12 = 106. At 0.05 significance, t-table gives 1.603. Since -5.25 < 1.603, accept H_0, which means there is no statistically significant bias found in the forecast.

4.6.4   Testing for the significance of regression coefficient.
Using the t-test, t-calculated is 14.91. With 106 degrees of freedom at 0.05 significance, t-table = 1.645. Since $14.91 > 1.645$, there is enough evidence to reject H_0 and H_1 accepted, meaning that the regression coefficient is statistically significantly different than zero.

Also, using the method of confidence interval test for β at a 95% confidence interval, we have

$b_1 \pm (t_{\alpha/2}, n-2-12) Se$, that is $0.321 \pm 0.075 = (0.246, 0.396)$

Since the confidence interval for β does not include the possibility zero it is statistically significant, $\beta \neq 0$, we assume that b cannot assume a value of 0 at the 5% level of significance and this suggests that the regression equation is a significant one, the slope of the coefficient is significantly different than zero.

Using the deseasonalized data the confidence interval are as shown in table 2.

Table 3: 95% confidence interval for β

95% Confidence Interval for β		
Model	Lower Bound	Upper Bound
Constant 1	1223373	374698
Period	0.246	0.396

The confidence intervals for the deseasonalized data are narrower than those of the raw data. In accordance with the fact that models with narrower confidence intervals are better models, then our model from the deseasonalized data is better than those of the raw data.

Considering the coefficient of determination the raw data gives $r^2 = 0.879$ while deseasonalized data gave $r^2 = 95.8$.

The model from desasonalized data is preferred because it has a high value of r^2 and a low value of residual standard error because it gives a low value for the sum of errors. Our result now shows that 95.8% of the variation in the volume of sales

can be explained by the relationship between volume of sales and the period, while 4.2% can not be explained or can only be explained by chance.

4.10   Forecast for the year 2017

Forecast was carried out for the year 2017 and from the forecast, the sum of squared error (SSE) = 3647655, mean absolute deviation (MAD) = 421.4938, mean error (ME) = 293.2621, mean absolute error (MAE) = 4214.938 and mean absolute percentage error (MAPE) = 68426 were obtained. These figures are minimal since the original data are in hundreds of thousands.

V.   DISCUSSION OF RESULTS

The result showed that deseasonalizing the data helps to reduce the fluctuations or smoothing out fluctuations that occur in the data. A plot of the residuals against time shows that one of the conditions that the variance of the error variable, r_e^2 be constant is satisfied. The error graph also shows that the series is trending and has seasonality in it, confirming that the series is not a random walk. The software used removed the trend by the method of twelfth difference and we had already removed seasonality, hence we have a white noise series. The use of t – test and confidence interval showed that the regression coefficients β were statistically significant.

5.1   Limitations

The details of the statistical tests for the bias of the model, significance of the regression coefficient, the Durbin-Watson test for autocorrelation, the t-test for autocorrelation function (ACF), test for correlation coefficient, test for outliers, the analysis of variance (ANOVA) table, table for confidence interval and table of forecasts for the year 2017 were not shown here due to space limitations.

5.2   Conclusion

In this work, we have seen that forecasting of the future sales of a particular product over a specific period of time is based on past performance of the product demand and satisfaction, weather, consumer spending patterns and market trends. Apart from marketing/sales, other areas where forecasting plays an important role includes inventory systems, financial planning and production planning. Indeed, managerial decisions seldom are made in the absence of some form of forecasting. Thus, a forecast is a basic tool to aid managerial decision making.

Two distinct quantitative techniques based on conventional statistical methods were used in this work; namely, time series analysis using the classical decomposition method and regression analysis. The time series techniques utilized data on sales volume collected over ten (10) years to forecast sales volume in a future period. This is adopted because, forecasting in general, is concerned with an analysis of past time series data in order to estimate one or more future values of the time series. In regression analysis, the variable to be forecast, sales volume (dependent variable) was expressed as a mathematical function of time period (independent variable). These two methods were used in conjunction with one another and the relationship between time and sales volume was developed as $\hat{Y} = 244957.77 + 0.3205608x.$ All the assumptions of linear regression model were satisfied and hence, the company can now use this model to predict their sales as this will give good prediction interval.

In conclusion, it has been shown that the use of deseasonalized data gives a better fit than using the raw data in the time series. The model developed accounted for 95.8% of the total variability with weather playing significant role in the sales volume as seen from the seasonal indexes.

REFERENCES

[1]. John, C. C., Satinder, K. M., & Donald. D. S. (2007). *How to choose the right forecasting technique*. Retrieved from https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique

[2]. Rob, J. H. & George, A. (2014). *Forecasting: principles and practice*. Retrieved from https://books.google.com.ng/books?hl=en&lr=&id=gDuRBAAAQBAJ&oi=fnd&pg=PA7&dq=forecasting:+principles+and+practice&ots=qKLobrz mhI&sig=M27j3VbnpkbqXi3kCoEqPFSw65o&redir_esc=y#v=onepage&q=forecasting%3A%20principles%20and%20practice&f=true

[3]. Shreyasi,   G.   (2016). *Sales   Forecasting: Meaning,   Factors,   Importance   and   Limitations*.   Retrieved   from http://www.yourarticlelibrary.com/sales/sales-forecasting-meaning-factors-importance-and-limitations/50997/

[4]. Susan, W. (2017). *Sales Forecasting*. Retrieved from https://www.thebalance.com/sales-forecasting-2948317

[5]. TrackMaven, (2017). *Sales Forecasting*. Retrieved from https://trackmaven.com/marketing-dictionary/sales-forecasting/