Journal of the Nigerian Association of Mathematical Physics
***Volume* 14 (May, 2009), pp 449 – 462**
© J. of NAMP

Modelling the Nigerian Defence Academy cadets' monthly sick-parade using time series classical decomposition method

[1]M. O. Oladejo and [2]L. U. Okafor
Department of Mathematics and Computer Science
Nigerian Defence Academy
Kaduna, Nigeria.

Abstract

The Nigerian Defence Academy (NDA) has no sick parade model that can enhance adequate planning and resource allocation. This research tackled this problem by developing a model that will assist in decision making policies.

Keywords
Classical decomposition, sick-parade, forecasting, fitting seasonal index, deseasonalized data, seasonalized data, moving average.

1.0 Introduction

Classical seasonal decomposition according to Mackenzi [7] is one of the oldest and most popular forecasting methods that can be used for strategic decisions and resource allocation. This method was first introduced in the mid 1950's by Shiskin J, with the Census I method to the present day X-11 variant of the Method II seasonal adjustment program.

The use of 12-month moving average in seasonal decomposition method using linear regression and seasonal index have attracted attention of researchers like: Black [3], Box and Jenkins [4], Fuller [5], Ghysels [6], Makridakis and Wheelwright [9], etc and all these used raw data to obtain the regression equation.

However, Andrew [1], Amir and Javayel [2], Stephen [11], Srivastava et al [12], etc, have shown that using deseasonalized data results in a better model selection than when raw data is used.

This research concentrates on the Census I method using the 12-month moving average and seasonally adjusted data (SAD), which is data obtained after dividing the original set of data by the adjusted or final seasonal factor (Si) for each month, where i = 1, 2, ..12 or i = 1, 2, 3, 4 for quarterly data. The statistical package for social sciences version 15 (SPSS15), has an inbuilt mechanism for obtaining SAD, S_i, Trend–cycle component (TC) and Irregular factor (It). The purpose of using SAD is to further reduce the fluctuations in the original data

After using the 12-month moving averages or 4-month moving averages for quarterly data this gave a better forecast for strategic decision and resource allocation.

[1]Corresponding author:
[1]Telephone: 08033430043
Telephone: 08036145954

1.1 Problem statement

There has been no model developed that can capture the forecast mechanism of cadets sick-parade at the NDA Medical Centre Kaduna since her inception in 1964.

1.2 Purpose of the study

The main purpose of this research is to develop a model that can be used to forecast the cadets monthly sick-parade considering the seasonal nature of the activities that these cadets do engage in from time to time so that accurate forecasts can be made about the number of cadets that may report sick in any given month. Consequently, the aim of this research work is to enhance the performance of the Academy Medical Centre through a better understanding of the demand for service at the centre. The model will enhance budgeting for efficient operation of the centre .

2.0 Methodology

2.1 Hypotheses of the research

Six hypotheses of the research are summarized as follows:

(1) H_0: $b_0 = 0$: The coefficients of the simple linear regression equation are zero.
H_1 : $b_0 \neq 0$: The coefficients of the simple linear regression equation are significantly different from zero.

(2) H_0: $\hat{e} = 0$: The mean error in the forecast is not statistically significantly different from zero.
H_1: $\hat{e} \neq 0$: The mean error in the forecast is statistically significantly t different from zero.

(3) H_0: $y = 0$: There is no strong linear relationship between the variables (time and cadets' sick parade)
H_1: $y \neq 0$: There is a statistically significant strong relationship between the two variables.

(4) H_0: ACF = 0 : The error e_t are auto correlated with first order scheme.
H_1: ACF $\neq$ 0 : The errors e_t not are auto correlated with first order scheme.

(5) H_0: CCF = 0 : The Cross-Correlation Coefficients (CCF(s)) are statistically significant.
H_1: CCF $\neq$ 0 : The Cross-Correlation Coefficients (CCF(s)) are not statistically significant.

(6) H_0: ACF : The auto-correlation functions (ACF$_{(s)}$) are statistically significant.
H_1: ACF $\neq$ 0 : The auto-correlation functions (ACF$_{(s)}$) are not statistically significant.

2.2 Significance of the study

This section is devoted to the justification of this research as a case study and the rational for using the classical seasonal decomposition Census I method.

(1) To isolate the components of the time series.
(2) To identify the seasonal patterns of each month.
(3) To develop a model from the sick-parade data that can be used to make predictions for the number of sick cadets to be expected every month for the next 2 – 3 years.
(4) To contribute to the growing literature in the field of forecasting using the 12-month moving averages of the classical seasonal decomposition method.

2.3 Assumptions of the Study

The following assumptions were based on those of Mark and David [8] and Tembe [11].

(1) The seasonal index for each month is stable.
(2) The model is predicated on the fact that present values can be predicted from its past history.
(3) Linear stochastic procedures apply to stationary time series.
(4) As in all forecasting the developed model may not give exact results always.
(5) The primary and secondary sources of data are accurate.
(6) The study is worth conducting for problem solving and decision making within operation research and seasonal decomposition computing.
(7) The software and hardware used in this study and research are the appropriate technology.
{8} This prototype as case study is accurate and appropriate fit.

2.4 Conceptual framework

Majority of the work done on classical seasonal decomposition using the 12-month moving averages for regression made use of the original date and the time variable: Black (3), Box and Jenkins [4], Fuller [5], Ghysels [6], Makridakis and Wheelwright [9]. The present work used deseasonalized data to derive the regression equation; this is in line with the views of Andrew [1], Stephen [11], and Srivastava et al. [12]. The statistical package used is the SPSS 15.

2.5 Materials and methods

Cadets' monthly sick parade from the Nigerian Defence Academy (Medical Centre) Kaduna for the months of January 2000 to the December 2006 is the source of data. This is shown in Table 2.1.

Table 2.1: Cadets monthly sick-parade (2000-2006).

	Year						
Month	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	124	100	129	94	93	100	137
Feb	111	105	97	99	87	129	160
Mar	175	158	180	184	195	182	107
Apr	125	230	100	120	142	236	210
May	190	180	283	184	152	300	150
Jun	185	204	159	221	144	215	136
Jul	151	178	143	202	110	145	73
Aug	220	254	231	231	297	376	169
Sep	200	298	194	200	209	145	169
Oct	250	300	350	210	206	244	179
Nov	170	123	167	172	236	250	205
Dec	90	80	87	82	130	160	156

4.6 Decomposition of time series components

There are two types of methods of decomposition of a time-series into the various components:

(a) $\hat{Y} = T{*}C{*}S{*}I$ (2.1)

The multiplicative method

(b) $\hat{Y} = T+C+S+I$ (2.2)

4.6.1 The additive method

In this work, the multiplicative method;

$$\hat{Y} = T{*}C{*}S{*}I$$

is used because it gives a better result than the additive method when mean error, Residual Standard Error (RSE) and other forecast accuracies are considered.

4.6.2 Moving Averages

The centered moving average assumes
$$X_t = \frac{\sum_{t=0}^{n-1/2} X_{t-1} + X_t + \sum_{t=1}^{n-1/2} X_{t-1}}{n} \qquad (2.3)$$

for n is odd.

$$X_t = \frac{\sum_{t=0}^{n/2} X_{t-1} + X_t + \sum_{t=1}^{n/2} X_{t-1}}{n} \qquad (2.4)$$

for n is even. But using SPSS 15, no formulae is needed as the package is already set for calculating centered moving averages. Table 2.2 shows the 12-month moving average for our data.

Table 2.2: 12-Month centred moving average

Month	Year						
	2000	2001	2002	2003	2004	2005	2006
Jan	_	171.5	181.75	172.83	159.17	196.08	189.06
Feb	_	173.75	178.83	177.75	151.5	199	183.08
Mar	_	176.58	176.92	177.75	157	205.58	165.83
Apr	_	184.75	168.25	178.25	157.75	204.33	163.75
May	_	188.92	172.25	166.58	157.42	207.5	158.33
Jun	_	185	176.08	167	162.75	208.67	154.58
Jul	165.92	184.17	176.67	166.58	166.75	211.7	_
Aug	163.92	186.58	173.75	166.5	167.33	214.25	_
Sep	163.42	185.92	173.92	165.5	170.83	216.83	_
Oct	162	187.75	174.25	166.42	169.75	210.58	_
Nov	170.75	176.92	175.92	168.25	177.83	208.17	_
Dec	169.92	185.5	167.67	165.58	190.17	195.67	_

2.7 Ratio of original series to moving average

Seasonal indexes (*S*) are calculated via the method of the ratio of the original data Y_t to the moving average (MA)

$$Y/MA = S.I*S; \qquad (2.5)$$

where *I* is the irregular factor and *SI* is seasonal index. Table 2.3 below shows the ratio of original series to moving average series %.

Table 2.3: Ratio of original series to original to moving average series %.

Month	2000	2001	2002	2003	2004	2005	2006
Jan	_	58.3	71	54.4	58.4	51	72.5
Feb	_	60.4	54.2	55.7	57.4	64.8	87.4
Mar	_	89.5	101.7	1034.5	1214.2	88.5	64.5
Apr	_	124..5	59.4	67.3	90	117	128.2
May	_	95.3	164.1	110.5	96.6	144.6	94.7
Jun	_	110.3	90.3	132.3	88.5	103	88
Jul	91	96.7	80.9	121.3	66	68.7	_
Aug	134.2	136.1	132.9	138.7	177.5	175.5	_
Sep	122.4	160.3	111.5	120.8	122.3	89.5	_
Oct	154.3	159.8	200.9	126.2	121.4	115.9	_
Nov	99.6	69.5	94.9	102.2	132.7	120.1	_
Dec	58.3	43.1	51.9	49.5	68.4	81.8	_

2.7.1 Seasonal index (*Si*)

Seasonal Index for each month of the year is obtained by calculating the average of each S.I for each month after eliminating extreme values in each month. Table 2.4 shows the monthly seasonal index *Si* as determined by the soft ware used.

Table 2.4: Adjusted seasonal indexes %

Month	Seasonal Index %
Jan	62.3
Feb	61.0
Mar	98.1
Apr	102.1

May	114.4
Jun	100.4
Jul	82.6
Aug	149.7
Sep	122.2
Oct	143.8
Nov	106.7
Dec	57

Table 2.4 is the monthly seasonal indexes deasonalizing the data or seasonally adjusted series (SAD). Having obtained the each monthly seasonal index, S_i, i = 1,2, …, 12, the data for each month's sick parade is now divided by its corresponding seasonal index S_i to get the deseasonalized value P_t for the month

$$P_t = Yt/SI \qquad (2.6)$$

t = 1, 2…,n.

Table 2.5 is the resulting values of the deseasonalized data as calculated by our software and its graph is shown in Figure 2.5.

Table 2.5: Deseasonalized data or seasonally adjusted series (SAS)

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	200.049	161.33	208.115	151.65	150.036	161.33	221.021
Feb	181.874	172.043	158.935	162.212	142.55	211.367	262.161
Mar	178.342	161.017	183.437	187.514	198.724	185.476	109.043
Apr	122.426	225.264	97.941	117.529	139.076	234.078	205.676
May	166.067	157.327	247.352	160.823	132.854	262.211	131.105
Jun	184.292	203.22	158.327	220.155	143.449	214.178	135.48
Jul	182.827	215.517	173.14	244.576	133.185	175.562	88.386
Aug	146.991	169.708	154.341	154.341	198.438	251.221	112.916
Sep	163.725	243.95	158.813	163.725	171.092	158.813	138.347
Oct	173.854	208.625	243.398	146.038	143.256	169.682	124.48
Nov	159.3	1154.259	156.489	161.175	221.146	234.265	192.098
Dec	157.819	140.284	152.559	143.791	227.961	280.568	273.553

2.7.1 Trend – Cycle

The Trend-Cycle (TC) for the series was calculated by the software and this has obviated the problem of obtaining the trend-cycle using a 3 × 3 moving average in order to overcome the loss of six values of the trend-cycle at the beginning and six values at the end of

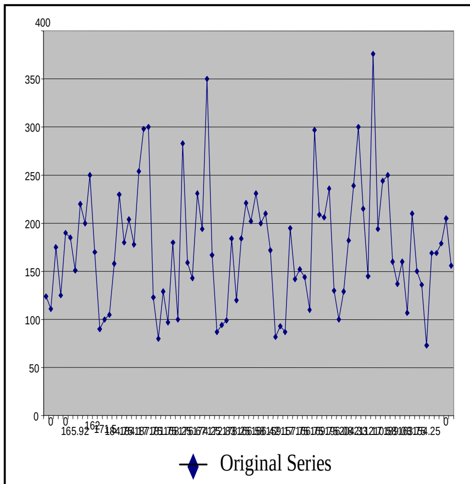

Figure 2.1: Original series

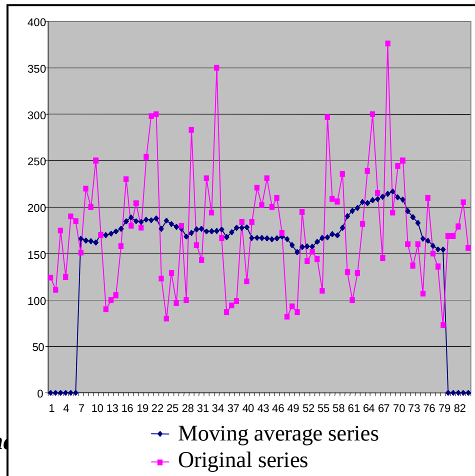

Figure 2.2: Moving average with original series

Table 2.6: Trend-Cycle values

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	196.258	162.67	169.053	158.721	143.632	196.585	232.426
Feb	186.755	171.545	166.459	159.388	156.449	198.861	214.762
Mar	167.749	177.369	168.837	159.45	160.257	207.873	179.437
Apr	158.029	187.527	163.636	159.07	151.82	224.795	166.107
May	163.645	189.498	179.033	176.658	143.947	227.131	141.451
Jun	168.898	194.48	174.271	193.681	144.438	222.598	129.335
Jul	171.204	199.298	172./339	200.807	154.142	208.723	114.451
Aug	165.803	204.434	169.857	182.8689	165.619	200.697	116.908
Sep	163.888	202.144	177.949	166.409	172.333	192.008	130.035
Oct	163.603	183.809	185.299	154.005	182.294	2002.999	157.867
Nov	162.923	166.`184	174.649	152.994	193.144	220.348	196.71
Dec	163.291	159.462	164.396	149.154	200.384	242.68	216.132

the series [i.e ($\frac{L}{12}$) values at the extreme ends of the series, where L is the seasonal length]. These missing values at the extreme ends are very important as they are needed to make forecasts at the end of period of (i.e. period 84 for our series] hence the need to get these extreme values replaced. Table 2.6 shows the Trend-cycle values for our series as calculated by SPSS 15.

2.7.2 Irregular patterns (I_t)

The SPSS 15 software isolated the irregular components I_t from the series.

$$Natural, I_t = Y/(T*C*S) \qquad (2.7)$$

Many authors Makridakis and Wheelwright [9], Srivastava et al [12]. etc concur that it is not easy mathematically to isolate the irregular component from a time series but SPSS 15 has an inbuilt mechanism that readily does that. Table 2.7 shows the irregular components of time series, and Figure 2.9 shows its graphical representation:

Table 2.7: Irregular components.

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	1.019	0.99921	1.231	0.955	0.977	0.821	0.951
Feb	0.974	1.003	0.955	1.017	0.911	1.063	1.221
Mar	1.063	0.908	1.086	1.176	1.24	0.892	0.608
Apr	0.775	1.202	0.599	0.739	0.916	1.041	1.238
May	1.015	0.83	1.382	0.91	0.923	1.154	0.927
Jun	1.091	1.045	0.909	1.137	0.993	0.962	1.048
Jul	1.068	1.081	1.005	1.218	0.864	0.841	0.771
Aug	0.887	0.83	0.909	0.844	1.198	1.252	0.966
Sep	0.999	1.207	0.892	0.984	0.993	0.827	1.064
Oct	1.063	1.135	1.314	0.948	0.786	0.836	0.789
Nov	0.978	0.694	0.896	1.053	1.145	1.063	0.977
Dec	0.972	0.88	0.928	0.964	1.138	1.156	1.266

This table contain values of I_t .

Trend Factors ($\hat{Y}_t$) would be estimated from the trend equations .

The trend equation, $Y = a + bx_t$ is obtained by the use of normal equations.

$$\sum P = axN + b\sum X \quad (2.8)$$
$$\sum XP = a\sum X + b\sum X2 \quad (2.9)$$

where, P stands for the seasonally adjusted data (SAD) or deseasonalized data. Here using SPSS 15, trend equation for a linear form was obtained as

$$T_1 = 170.283 + 0.148t \quad (2.10)$$

Table 2.8 shows the complete table for the deterministic model and Figure 2.6 is a graphical illustration of trend and fitted values.

Table 2.8: Trend values using $\hat{Y}$ = (170.283 +.148*t)

Month	**Year**						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	170.4314	172.2074	173.9834	175.7594	177.5354	179.3114	181.0874
Feb	170.5794	172.3554	174.1314	175.9074	177.6834	179.4594	181.2354
Mar	170.7274	172.5034	174.2794	176.0554	177.8314	179.6074	181.3834
Apr	170.8754	172.6514	174.4274	176.2034	177.9794	179.7554	181.5314
May	171.0234	172.7994	174.5754	176.3514	178.1274	179.9034	181.6794
Jun	171.1714	172.9474	174.7234	176.4994	178.2754	180.0514	181.8274
Jul	171.3194	173.0954	174.8714	176.6474	178.4234	180.1994	181.9754
Aug	171.4674	173.2434	175.0194	176.7954	178.5714	180.3474	182.1234
Sep	171.6154	173.3914	175.1674	176.9434	178.7194	180.4954	182.2714
Oct	171.7634	173.5394	175.3154	177.0914	178.8674	180.6434	182.4194
Nov	171.9114	173.6874	175.4634	177.2394	179.0154	180.7914	182.5674
Dec	172.0594	173.8354	175.6114	177.3874	179.1634	180.9394	182.7154

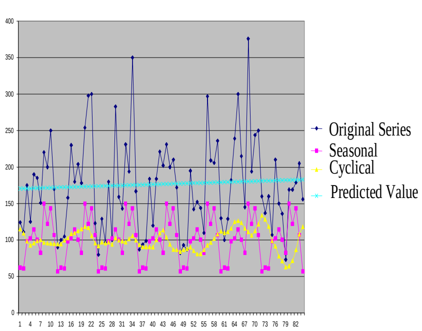

Figure 2.7: Original, Seasonal and Cyclical with Trend

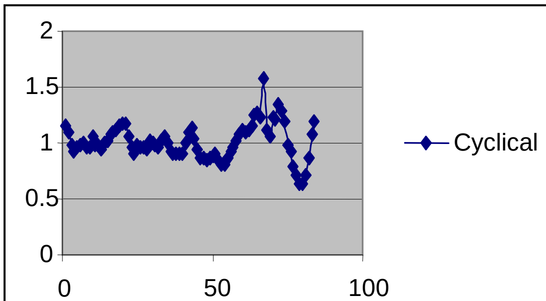

Figure 2.8: Cyclical components

2.7.3 Cyclical component (C)

The cyclical component (C) was computed from the relation

$$\frac{Tc}{T} = \mathrm{c} \quad (2.11)$$

Using values of Tc and T from Tables 2.7 and 2.8 respectively to yield the values shown in Table 2.9 contains the cyclical component and Figure 2.8 shows its graph. The fitted value or trend seasonal factor ($\hat{Y}$) is obtained by multiplying the Trend value (T) and the seasonal index S_i for any particular month to get:

$$\hat{Y}_t = T_t{*}S_i \quad (2.12)$$
$$= (a + b{*}t){*}S_i \quad (2.13)$$
$$\hat{Y}_t = (170.283{+}0.148{*}t){*}S_i \quad (2.14)$$
where $T_t = (a + b{*}t),\ T = 1, 2,\ldots,n,\ i = 1, 2, \ldots.\ 120.$ (2.15)

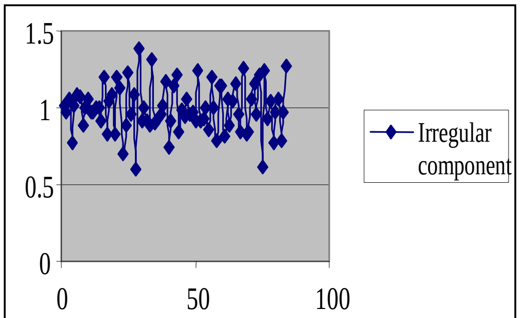

Figure 2.9: Irregular components

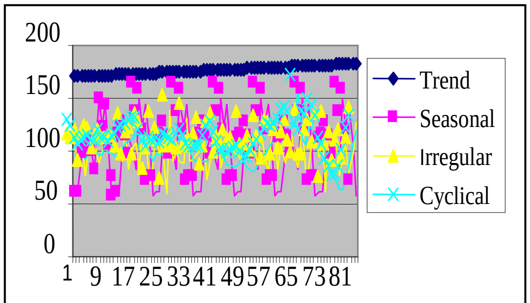

Figure 2.10: Component of a time series

Table 2.9: Cyclical components

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	1.151537	0.945	0.972	0.903	0.865	1.096	1.283
Feb	1.094827	0.995	0.956	0.906	0.88	1.108	1.185
Mar	0.982554	1.028	0.969	0.905	0.901	1.157	0.989
Apr	0.92482	1.086	0.938	0.902	0.853	1.251	0.915
May	0.956856	1.097	1.026	1.002	0.808	1.262	0.779
Jun	0.986	1.124	0.997	1.097	0.81	1.236	0.711
Jul	0.999	1.151	0.986	1.137	0.864	1.58	0.63
Aug	0.967	1.18	0.97	1.034	0.927	1.113	0.642
Sep	0.955	1.166	1.016	0.94	0.964	1.064	0.713
Oct	1.063	1.059	1.057	0.87	1.019	1.234	0.865
Nov	0.978	0.957	0.995	0.863	1.079	1.219	1.077
Dec	0.972	0.9127	0.917	0.841	1.118	1.341	1.183

Table 2.9 shows the fitted values of $\hat{Y}$ while Figure 2.6 shows fitted and trend values
Fitted Values using $\hat{Y} = (170.283{+}0.148{*}t){*}S_i$

2.8 Fitted errors

The fitted error e_t for time t is given as

$$e_t = y_t - \hat{y}_t \quad (2.16)$$

where y_t is actual value for month t, $\hat{Y}_t$ is a fitted value for month t. Table 2.10 shows the fitted errors while its graph is shown in Figure 2.5. Table 2.10 has values of filled errors.

Table 2.10: Fitted values using $\hat{Y} = (170.283 + 0.148{*}t){*}S_i$

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	106.178762	107.28521	108.391658	109.59031	110.696758	111.803206	112.909654
Feb	104.053434	105.136794	106.220154	107.393794	108.477154	109.560514	110.643874
Mar	167.483579	169.225835	170.968091	172.855535	174.597791	176.340047	178.082303

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Apr	174.463783	176.277079	178.090375	180.054779	181.868075	183.681371	185.494667
May	195.65077	197.682514	199.714258	201.915314	203.947058	205.978802	208.010546
Jun	171.856086	173.63919	175.422294	177.35399	179.137094	180.920198	182.703302
Jul	147.677323	149.208235	150.739147	152.397635	153.928547	155.459459	156.990371
Aug	256.686698	259.34537	262.004042	264.88427	267.542942	270.201614	272.860286
Sept	209.714019	211.884291	214.054563	216.405691	218.575963	220.746235	222.916507
Oct	246.995769	249.549657	252.103545	254.870257	257.424145	259.978033	262.531921
Nov	183.429464	185.324456	187.219448	189.272356	191.167348	193.06234	194.957332
Dec	98.073858	99.086178	100.098498	101.195178	102.207498	103.219818	104.232138

Table 2.11: Fitted errors

Month	Year						
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Jan	18.33253	-6.76896	9.030731	-13.0995	-17.0734	-11.1749	24.72364
Feb	6.946566	-0.13716	-78.0917	-14.9719	-21.3883	19.528	49.44427
Mar	7.516323	-11.2265	83.28426	11.28789	20.54504	5.802198	-70.9406
Apr	-49.4639	53.72221	-16.4237	-59.9056	-39.7195	55.46667	24.65277
May	-5.651	-17.6834	-1.44493	-17.7482	-51.7806	94.18696	-57.8455
Jun	13.14371	30.36001	-31.0063	43.79259	-34.9911	34.22528	-46.5584
Jul	9.489928	35.02246	-20.0564	56.0876	-37.3799	-3.84735	-77.3148
Aug	-36.6871	-5.34672	97.89415	-33.6659	29.67472	106.0152	-103.644
Sep	-9.71451	86.11461	-20.2212	-16.2274	-9.39841	-26.5694	-53.7404
Oct	3.003656	50.4489	-13.0995	-44.6606	-51.2152	-15.77	-83.3247
Nov	-13.43	-62.3255	-14.9719	-17.1168	44.98758	57.09195	10.19632
Dec	-8.07414	-19.0868	-13.0995	-19.1121	27.87527	56.8626	51.84994

3.0 Analysis ACF/PACF

Autocorrelation function (ACF), partial auto correlation function (PACF) and their standard errors (*Si*) were computed using SPSS 15. The software used produced ACF(*k*) and PACF(*k*) and their standard errors as shown below – Table 3.1 and its graph is shown in Figure 3.1.

3.1 Cross-Correlation Coefficients CCF(*k*)

Cross-correlations which measures the strength of relationship between Y_t and X_{t-k} can be calculated from the relationship. Results produced by the software used is as shown in Table 3.2 and its graph is shown in Figure 3.2 including the standard errors (S_e). The CCF(*k*) is given by

$$r\left(Y_t X_{t-k}\right) = \frac{Cov\left(Y_t X_{t-k}\right)}{Sy_t Sx_{t-k}},$$

where Sy_t , Sx_{t-k} are the standard errors of y_t and x_{t-k} respectively. By definition correlation coefficient is given by

$$r_{xy} = \frac{Cov(X,Y)}{S_x S_y},$$

and covariance of XY by

$$Cov(X,Y) = \frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{(n-1)}$$

Table 3.1: Residual ACF, PACF, and their S_e summary

ACF	PACF			
Lag	***ACF***	S_e	***PACF***	S_e
Lag 1	0.066	0.109	0.066	0.109
Lag 2	0.059	0.11	0.055	0.109
Lag 3	0.143	0.11	0.137	0.109
Lag 4	0.037	0.112	0.018	0.109
Lag 5	0.141	0.112	0.126	0.109
Lag 6	-0.104	0.114	0.146	0.109
Lag 7	-0.15	0.115	0.163	0.109
Lag 8	0.137	0.118	0.137	0.109
Lag 9	0.064	0.12	0.101	0.109
Lag 10	-0.187	0.12	-0.2	0.109
Lag 11	-0.217	0.123	-0.235	0.109
Lag 12	-0.058	0.1280.128	-0.006	0.109
Lag 13	0.026	0.128	0.064	0.190
Lag 14	-0.161	0.128	-0.13	0.190
Lag 15	-0.147	0.131	-0.037	0.190
Lag 16	-0.114	0.133	-0.081	0.190
Lag 17	-0.029	0.134	-0.111	0.190
Lag 18	-0.058	0.134	-0.068	0.190
Lag 19	-0.183	0.134	-0.016	0.190
Lag 20	-0.018	0.137	0.038	0.190
Lag 21	-0.022	0.137	-0.151	0.190
Lag 22	0.111	0.137	0.095	0.190
Lag 23	-0.001	0.138	0.045	0.190
Lag 24	-0.064	0.138	-0.09	0.190

Table 3.2: Cross correlation

Lag	**Cross Correlation**	**Standard Error**
	$CCF_{(k)}$	S_e
-7	-0.01	0.125
-6	0.006	0.124
-5	-0.01	0.123
-4	-0.039	0.122
-3	0.137	0.121
-2	-0.099	0.12
-1	-0.033	0.12
0	-0.058	0.119
1	-0.036	0.12
2	-0.026	0.12
3	-0.018	0.121
4	-0.022	0.122
5	-0.018	0.123
6	-0.013	0.124
7	-0.015	0.125

where, $\overline{X}$ and $\overline{Y}$ are the means and, S_x and S_Y are standard errors of estimates of X and Y respectively. Table 3.2 is the Cross-Correlation Coefficients, CCK(t) The graph is represented in Figure 3.2.

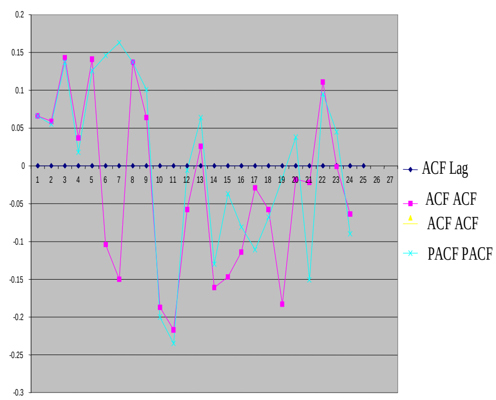

Figure 3.1: AFC and PACF

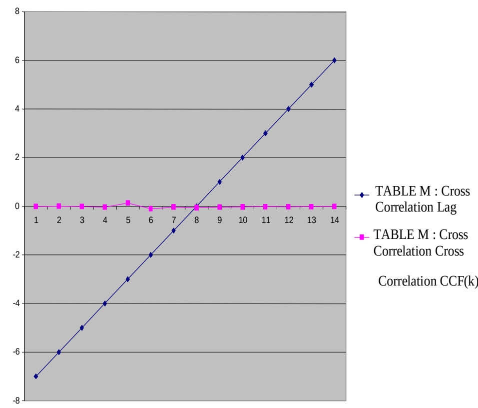

Figure 3.2: Cross Correlation

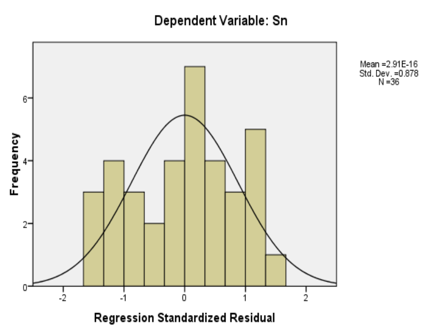

Figure 3.3: Histogram of residuals

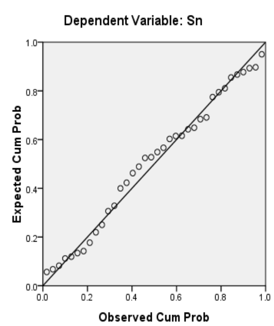

Figure 3.4: Normal probability plot

3.2 Forecast for 2007 or out of sample FORECAST

Using the model $\hat{Y}_t = [170.283 + 0.148*t]*S_i$, values were computed as shown in Table 3.1. Using Theil's [14] inequality coefficient

$$U^2 = \frac{\sum (P_i - A_i)^2 / n}{\sum A_i^2 / n},$$

were also computed , where P_i = forecast change in the dependent variable, A_i = actual change in the dependent variable.

3.3 Analysis of variance ANOVA

Using the trend equation without the seasonal index, $\hat{Y}_t = 170.283 + 0.148*t$, the software used produced the ANOVA table shown below.

Table 3.3 ANOVA

Model	Sum of Squares	*df*	Mean Square	*F*	Sig
Regression	945.98	1	945.98	0.231	0.632
Residual	336071.59	82	4098.434		
Total	337017.57	83			

But when the trend equation is used with the seasonal index (S_i) , since $F(0.231) < F5\%(1.82)$ = accept H_0 that model is a good fit. $\hat{Y}_t = [170.283 + 0.148*t]*S_i$.

4.0 Discussion of the results

The result showed that deseasonalizing the data helps to reduce the fluctuations that occur in the time series. A plot of residuals (e_t) against time (t) shows that there is a random behaviour of the residual indicating that there is no Heterosadasticity, thus confirming that there is no model inadequacy. The plot is shown in Figure 2.6. A plot of the histogram of the residuals and the shape of this histogram is in Figure 3.3 appears to show that the residuals are normally distributed. Besides, the Normal probability plot of the distribution of the errors indicates that by the way the errors are aligned along the straight line in Figure 3.4, the residuals are normally distributed a further confirmation of the assumptions of the simple linear regression. The error terms are independent as they are not correlated and this is a further prove that there is no autocorrelation in the errors, the fact that was conformed by the use of Durbin-Watson *t*-test.

The Durbin-Watson statistics confirms that ACF(k) or at any lag for the series is statistically not significant and this denotes white noise.

Z-test was used to detect outliers and none was found, a further show that not much mistake was made in recording the data points included in this sample.

4.1 Findings

The following findings resulted from this work

(1) Deseasonalizing the data reduces fluctuations by assigning more weights to originally low data values. This is shown in Table 2.5.

(2) The cyclical factor is almost constant for each year but the seasonal factor changes monthly. This is shown on Tables T and D in that order.

(3) Trend equation obtained using deseasonalized data gives a lower or minimum sum of squares of errors or SSE than when raw data is used to obtain the trend equation. See ANOVA Table 3.3.

(4) The Linear model gives a better fit statistics than any other model.

(5) Confidence intervals are narrower than prediction intervals for estimation.

(6) Seasonality/seasonal index affect the Residual Standard Error and other errors in a series and reduce the coefficient of determination.

(7) Irregular Components have no identifiable pattern. See Table 2.9.

(8) It is not in every case that classical decomposition method yields residuals that are extremely auto correlated in time series data.

Acknowledgement

The authors wish to acknowledge the constructive criticisms of Professors R. O. Ayeni and G. C. Ovuworie which helped to improve this work a great deal.

References

[1] Andrew. F.Siegel: Practical Business Statistics (2002),McGraw-Hill Iwrin.
[2] Amir A. and Jayavel S.: Complete Business Statistics (2002), McGraw-Hill Irwin.
[3] Black Ken: Contemporary Decision Making (2001), West Publishing Company.

[4] Box,G.E.P. and G.M Jenkins: Time Series Analysis, Forecasting and Control (1976, 1984, 1989), San Francisco, Holden-Day.
[5] Fuller W.A.: Introduction to Statistical Time Series (1970), New York.

[6] Ghysels.E.(1994): On the Economics and Econometrics of Seasonality in C.A. Sims9 (ed) Advances in Econometrics, Sixth World Congress of Econometrics Society, Cambridge; Cambridge University Press.

[7] Mackenzie. E.: General Exponential Smoothing and the Requirement ARMA. Journal of Forecasting 3 (1984) pp. 333-344.

[8] Mark L. B. & David M. M. L. Basic Business Statistics, Concepts and Application, Prentice Hall Inc, Engle Wood Cliffs (1979)

[9] Makridakis S.C. and Wheelwright S.C.: Forecasting Methods for Management (1989), John Wiley & Sons Inc.

[10] Nie N., C. Hull, J Jenkins, K. Steinbrenner and D. Bert: SPSS 15 Statistical Package for Social Sciences, New York: (McGraw Hill) (2006)

[11] Stephen A. DeLurgio: Forecasting Principles and Applications (1998), Iwrin McGraw -Hill . Cambridge

[12] Srivastava U.K., G.V. Shenoy and S.C.Sharma: Quantitative Technique for Managerial Decisions (2008), New Age International Publishers

[13] Tembe Elias .O.A.: Artificial Intelligence (AI) using Natural Language (NPL) with Operations Research (OR), University of Dubug (2005), etmbe @ obq.Ed.

[14] Theil H.: Applied Economic Forecasting (1966), North Holland.