Regression Analysis and Seasonal Influence Analysis of Production Yield in Finoplastika Industry

[1]Ezeliora Chukwuemeka Daniel, [2]Onyechi Pious Chukwukelue, [3]Chinwuko Emmanuel Chuka and [4]Ezeanyim Okeychukwu Chiedu

[1,2,3,4]Department of Industrial and Production Engineering, Nnamdi Azikiwe University Awka, Anambra State, Nigeria

[1]cezeliora@gmail.com, [2]pconyechi@yahoo.co.uk, [3]chinwukoemmanuelc@yahoo.com

Abstract– **This research work was used to address the problem of Analysis of the forecasting of plastic yield in Finoplastika manufacturing industry. Data on production yield were collected from the industry covering a period of three years. Forecasting based time series technique was applied to understand the seasonal influence of their production system. However, there is always a sharp increase of product total in the month of April and a sharp fall in the month of October and November concurrently for over three years. It can also observe that the least product total indices are usually in the month of Septembers as it is express graphically. Furthermore, from the analysis of variance, when product total is dependent and p1-p8 are the independent, the coefficient of determination for the model obtained is r^2= 98.7% which means that 98.7% is the proportion of variation in the dependent variable (month) that is explained by the variation in the independent variable (product total). Hence the model is fit for predicting the product total, since the coefficient of determination shows a strong relationship. The model was therefore used to suggest optimum monthly production output for different product types investigated. This will prevent the incident of under producing or over producing as identified.**

Keywords– **Regression, Seasonal Influence Forecasting, Production and Data**

I. INTRODUCTION

Finoplastika manufacturing industry limited is a company producing plastics in different shapes and sizes. They produce both injection and extrusion plastics. Injection plastics are domestic plastics (e.g. bucket, cup, spoon, etc.), plastic furniture (e.g. chairs, tables, etc) and industrial plastics (e.g. Toilet cover, laptop cover, television cover, etc. Over the years, the company has made substantial progress. However there is no formalized way of determining what quantity of different products to be produced over any given period of time. This leads to inconsistencies in planning and production. To avoid any of these problems during production, there is a need to optimize in advance the quantity of products, in other to support decision making regarding quantity of the plastic products necessary for production in every month.

The objective of this report is to develop a time series technique and regression analysis in order:

- ✓ To extract the seasonal index generated for the product types.
- ✓ A regression analysis was also used to understand change in the independent variable that was explained in the dependent variable.
- ✓ To help us study the various components, that plays a major role in the decision making and market strategy.
- ✓ To optimize the Production planning system in the manufacturing industry.
- ✓ To make recommendation to the company based on the research findings.

A) Time -Series Analysis and Forecasting

A time series is a sequence of observations obtained through measurements often recorded at equally spaced intervals. Often, time series data have characteristics that facilitate forecasting. These include seasonality, underlying trends, and relationships with past observations or other causal variables. Analysts can improve time series forecasts if they understand the nature of these components and identify the model that will best exploit the data's characteristics.

The purpose of this chapter is to provide a synopsis of time series analysis and forecasting. The first section discusses the characteristics of time series data. It reviews the common components useful in creating effective forecasts such as trend, seasonality, cyclical behavior, and irregular fluctuations. The chapter concludes with an introduction to time series forecasting and an overview of the forecasting model development process (Douglas I. Feiring, 1990).

Characteristics of Time-Series Data: time series is a "collection of observations made sequentially in time" (Chatfield, 1996). Examples are records of local daily rainfall levels, the quarterly U.S. Gross Domestic Product, and the monthly Marine Corps personnel strength for a particular rank and MOS. Time series analysis provides tools for choosing a representative model and producing forecasts. There are two kinds of time series data:

• Continuous, where the data contain an observation at every instant of time, e.g., seismic activity recorded on a seismogram.

• Discrete, where the data contain observations taken at intervals, e.g. monthly crime figures.

Unless the data are purely random, observations in a time series are normally correlated and successive observations may be partly determined by past values (Chatfield, 1996). For example, the meteorological factors that affect the temperature on any given day are likely to exert some influence on the following day's weather. Thus, historical temperature observations are beneficial in forecasting future temperatures.

A time series is deterministic if it contains no random or probabilistic characteristics but proceeds in a fixed, predictable fashion (Chatfield, 1996). An example of a deterministic time series would be the data collected while conducting a classical physics experiment such as one demonstrating Newton's law of motion (Gujarati, 2003). More applicable to econometric applications are stochastic time series. Stochastic variables have indeterminate or random aspects. Although the values of individual observations cannot be predicted exactly, measuring the distribution of the observations may follow a predictable pattern. Statistical models can describe these patterns. These models assume that observations vary randomly about an underlying mean value that is a function of time. Time series data can also be characterized by one or more behavioral components: trend, seasonality, cyclical behavior, and random noise.

Trend Component: Trend is the general drift or tendency observed in a set of data over time. It is the underlying direction (an upward or downward tendency) and the degree of change in an observation set when consideration has been made for other components. Graphing a time series can be a useful and simple method of identifying the trend of a particular data set. This indicates an upward trend of the U.S. Gross Domestic Product over a ten year span. Analysts can also discern trends by dividing the data set into a number of ranges, and calculating the mean for each span. A consistent increase or decrease in the mean for the successive ranges indicates trend.

Trends in business or economic series may be due to a growth or contraction process. Trends in service manpower levels may be attributed to external economic factors or shifts in policy due to technical innovation, downsizing, or an increased or decreased operational requirement for certain occupational specialties.

Seasonal Component: In time series data, the seasonal component is the element of variation in a data set that is dependent on the time of year. Seasonality is quite common in econometric time series. It is less common in engineering and scientific data. This component recurs annually, with possible variations in amplitude. Seasonality is attributable to the change of seasons and/or the timing of such events as holidays or the start or completion of the school term. For example, the cost of fresh produce, retail sales levels, average daily rainfall amounts, and unemployment figures all demonstrate seasonal variation.

Incorporating seasonality in a forecast is useful when the time series has a discernible seasonal component. When the data contain a seasonal effect, it is useful to separate the seasonality from the other components in the time series. This enables the analyst to estimate and account for seasonal patterns.

Cyclical Component: Cyclical behavior describes any non-seasonal component that oscillates in a recognizable pattern. The 11-year sunspot cycle has been long recognized as naturally occurring cyclical activity. More ambiguous is the 5 to 7 year business cycle that a number of economists hypothesize influence global economic activity. If the data include a discernible cyclical component, the time series should span enough cycles to accurately model and forecast its effects (Yaffee, 2000). Cyclical behavior in which the oscillations extend over a very long period (such as 20 years) can often be accurately modeled as a trend for short-term forecasts (Chatfield, 1996).

General linear model: The general linear model (GLM) is a statistical linear model. It may be written as (Christensen, 2002). $\mathbf{Y} = \mathbf{XB} + \mathbf{U}$, where **Y** is a matrix with series of multivariate measurements, **X** is a matrix that might be a design matrix, **B** is a matrix containing parameters that are usually to be estimated and **U** is a matrix containing errors or noise. The errors are usually assumed to follow a multivariate normal distribution. If the errors do not follow a multivariate normal distribution, generalized linear models may be used to relax assumptions about **Y** and **U**.

The general linear model incorporates a number of different statistical models: ANOVA, ANCOVA, MANOVA, MANCOVA, ordinary linear regression, t-test and F-test. The general linear model is a generalization of multiple linear regression model to the case of more than one dependent variable. If **Y**, **B**, and **U** were column vectors, the matrix equation above would represent multiple linear regression.

Hypothesis tests with the general linear model can be made in two ways: multivariate or as several independent univariate tests. In multivariate tests the columns of **Y** are tested together, whereas in univariate tests the columns of **Y** are tested independently, i.e., as multiple univariate tests with the same design matrix (Wichura, 2006).

B) Linear Regression Models

In statistics, linear regression models often take the form of something like this:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \varepsilon$$

Here a response variable y is modeled as a combination of constant, linear, interaction, and quadratic terms formed from two predictor variables x_1 and x_2. Uncontrolled factors and experimental errors are modeled by ε. Given data on x_1, x_2, and y, regression estimates the model parameters β_j (j = 1... 5).

More general linear regression models represent the relationship between a continuous response y and a continuous or categorical predictor **x** in the form:

$$y = \beta_1 f_1(\boldsymbol{x}) + \ldots + \beta_p f_p(\boldsymbol{x}) + \varepsilon$$

The response is modeled as a linear combination of (not necessarily linear) functions of the predictor, plus a random error ε. The expressions $f_j(\mathbf{x})$ (j = 1, ..., p) are the terms of the model. The β_j (j = 1... p) are the coefficients. Errors ε are assumed to be uncorrelated and distributed with mean 0 and constant (but unknown) variance.

Whether or not the predictor **x** is a vector of predictor variables, multivariate regression refers to the case where the response **y** = (y_1, ..., y_M) is a vector of M response variables. See Multivariate Regression for more on multivariate regression models (Rawlings, 1998).

C) Multiple Linear Regression

Multiple linear regression, is a generalization of linear regression (Mardia, 1979), by considering more than one independent variable, and a specific case of general linear models formed by restricting the number of dependent variables to one. The basic model for linear regression is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \epsilon_i.$$

In the formula above we consider n observations of one dependent variable and p independent variables. Thus, Y_i is the i^{th} observation of the dependent variable, X_{ij} is i^{th} observation of the j^{th} independent variable, j = 1, 2, ..., p. The values β_j represent parameters to be estimated, and ε_i is the i^{th} independent identically distributed normal error (Friston, 1995).

II. RESEARCH METHOD USED

The research method used in this work is a quantitative research approach. The data gathered were the daily record of plastic pipes production over the month for three years. The research method emphasis detailed analysis of regression analysis to show the change in the independent that was explained in the dependent. Furthermore, time series technique was also use to understand in details the analysis of the seasonal influence that shows and observes the seasonal influence of the data on monthly basis for a period of three years respectively. The use of Minitab tool was applied for the development of various analyses and the achievement of the results.

Company Data Presentation: the company production quantity of the data is shown in Table 1:

Table 1: Presentation of 2009-2011 Monthly Data on Quantity of finished products of Finoplastika industries ltd, Nigeria

Year	*Month*	*PT*	*p1*	*p2*	*p3*	*p4*	*p5*	*p6*	*p7*	*p8*
2009	Jan	50488	16526	3860	9618	15571	0	4493	420	0
	Feb	76031	29250	40	14773	10680	390	18718	2180	0
	Mar	74010	26666	9960	16571	11280	453	6740	2340	0
	Apr	123767	52029	10315	32339	11660	0	12940	4484	0
	May	70704	14160	17241	10788	14540	0	9560	4415	0
	Jun	47610	23087	2340	878	6146	0	8475	6684	0
	Jul	77654	29890	26785	15885	1140	0	3040	700	214
	Aug	61053	17981	20280	9062	1540	0	12140	50	0
	Sep	13538	3248	0	7570	2260	0	0	460	0
	Oct	21476	7045	7530	2611	2120	0	454	1716	0
	Nov	40561	16014	3768	5883	2980	0	6002	5914	0
	Dec	4871	3171	280	0	1160	0	260	0	0
2010	Jan	28462	7113	6311	8445	4693	0	1360	540	0
	Feb	16154	7284	0	4595	1760	0	390	2125	0
	Mar	70844	22119	24975	9535	7295	560	6340	20	0
	Apr	64666	21134	0	18843	15480	0	2930	6279	0
	May	46107	18848	4545	4497	4180	0	7760	6367	0
	Jun	49058	22172	4920	14296	2589	0	3733	1205	143
	Jul	33287	8767	13790	2351	2278	0	3040	3061	0
	Aug	37849	14790	1740	10885	3900	0	2080	4454	0
	Sep	29459	11975	0	15023	0	0	1360	1101	0
	Oct	25738	5518	2245	5049	1740	583	3760	6843	0
	Nov	35740	17532	1830	9948	3640	60	2730	0	0
	Dec	60455	18452	360	9489	8120	0	280	23754	0
2011	Jan	53480	22225	160	20184	3724	651	2860	2860	816
	Feb	31729	14123	2140	4721	5620	0	2340	1408	1377
	Mar	42625	14502	2200	11137	6680	262	7600	0	244
	Apr	36237	16014	910	1970	8880	0	4560	3497	406
	May	63066	24134	1062	21265	7720	255	8060	570	0
	Jun	60997	29097	5300	20838	16160	607	7750	0	0
	Jul	61892	16981	17170	6210	7500	605	10822	2604	0
	Aug	58988	17298	7545	11877	11420	733	6020	4095	0
	Sep	41820	5617	20085	2421	5980	277	6820	620	0
	Oct	69547	20631	5960	16326	6220	604	14310	5496	0
	Nov	11616	4391	0	1720	4173	52	1280	0	0
	Dec	29053	11909	1760	1706	6610	558	6510	0	0

A) Method of data analysis

In the method of data analysis, some group of data were analyzed by using seasonal influence and regression analysis to show the change in the independent that was explained in the dependent in the manufacturing industry. The use of minitab tool was made to test for the various analyses.

Seasonal Analysis on quarterly production yield of each finished products

Estimation of the seasonal influence on the monthly production of finished products, the table below shows the seasonal analysis summary result using the Minitab package.

Data analysis for seasonal index on quarterly production yield of each finished products

Data pT
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	1.00873
2	0.588212
3	1.48163
4	1.28332
5	1.29416
6	1.30730
7	1.14819
8	1.06664
9	0.495206
10	0.567041
11	0.952951
12	0.806607

Data p1
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	0.930916
2	0.703043
3	1.28979
4	1.26785
5	1.41338
6	1.68228
7	1.06406
8	0.960893
9	0.400745
10	0.406309
11	1.12364
12	0.687092

Data p2
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	0.493301
2	0.381814
3	2.78938
4	0.098193
5	0.613353
6	1.11539
7	3.37623
8	1.52443
9	0
10	0.888335
11	0.622810
12	0.096769

Data p3
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	1.42607
2	0.514421
3	1.11518
4	1.23425
5	1.22302
6	1.78771
7	0.873396
8	0.971644
9	0.11938
10	0.409569
11	0.888709
12	0.436636

Data p4
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	0.985865
2	0.707518
3	1.48240
4	2.74965
5	1.11492
6	1.48113
7	0.368747
8	0.603945
9	0.249465
10	0.442695
11	0.797085
12	1.01658

Data p5
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	1.49544
2	0
3	6.66978
4	0
5	0.393478
6	0.877078
7	0
8	0
9	0
10	2.34150
11	0.222727
12	0

Data p6
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	0.586260
2	0.345583
3	2.04988
4	0.969340
5	2.12714
6	1.38491
7	0.808778
8	1.51831
9	0.028778
10	0.664746
11	1.23677
12	0.079507

Data p7
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	0.561626
2	0.702198
3	0.004284
4	1.76866
5	1.35485
6	0.190974
7	0.539406
8	0.548426
9	0.003648
10	1.27657
11	1.37623
12	3.43013

Data p8
Length 36.0000
NMissing 0

Seasonal Indices

Period	Index
1	1.14808
2	1.93739
3	0.343299
4	0.571228
5	0
6	4.00000
7	4.00000
8	0
9	0
10	0
11	0
12	0

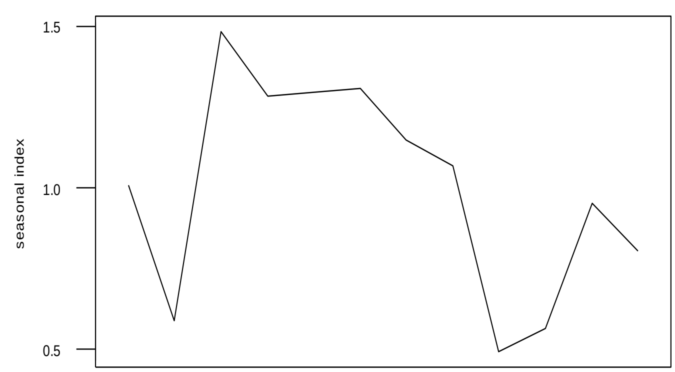

Figure 1: Seasonal Indices Plot of total production yield of finished products (P_T)

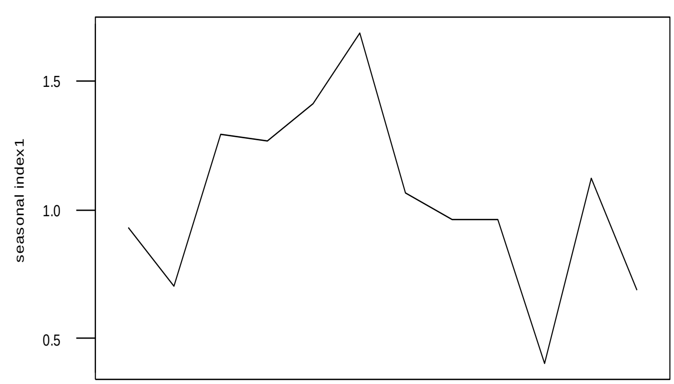

Figure 2: Seasonal Indices Plot of total production yield of finished products (P_1)

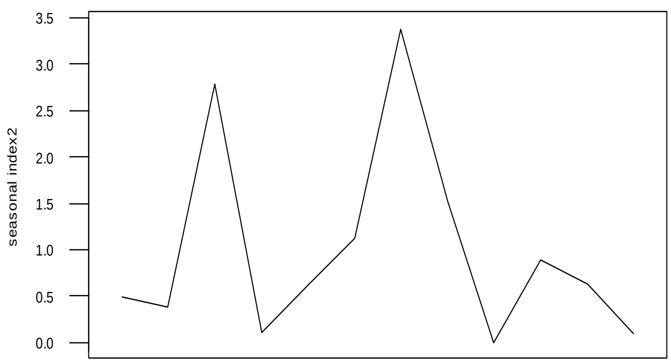

Figure 3: Seasonal Indices Plot of total production yield of finished products (P_2)

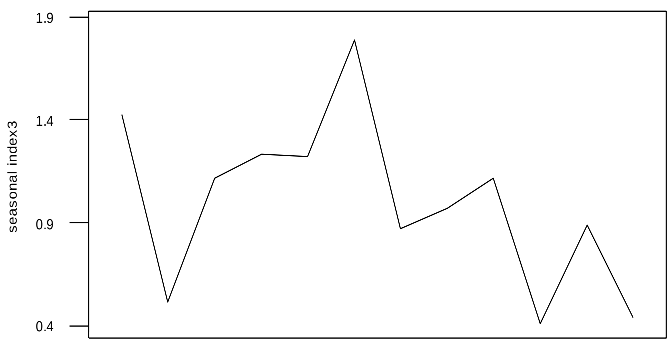

Figure 4; Seasonal Indices Plot of total production yield of finished products (P_3)

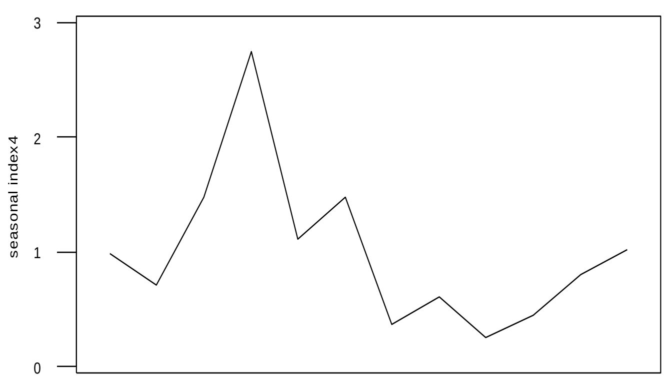

Figure 5: Seasonal Indices Plot of total production yield of finished products (P_4)

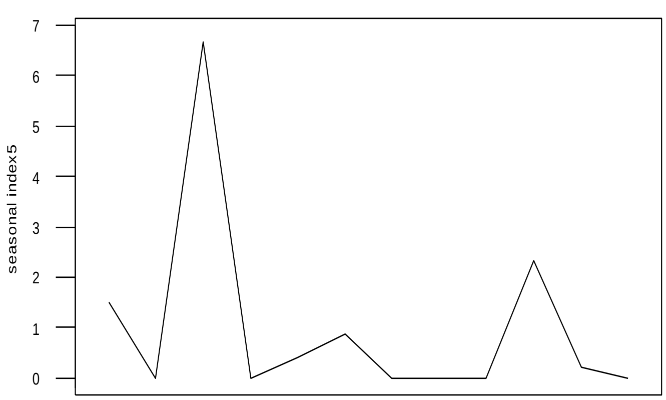

Figure 6: Seasonal Indices Plot of total production yield of finished products (P_5)

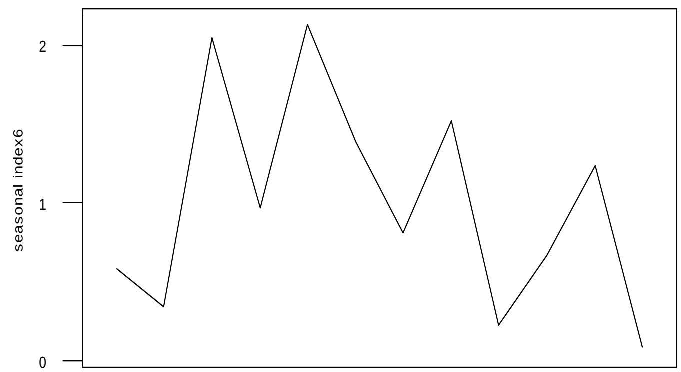

Figure 7: Seasonal Indices Plot of total production yield of finished products (P_6)

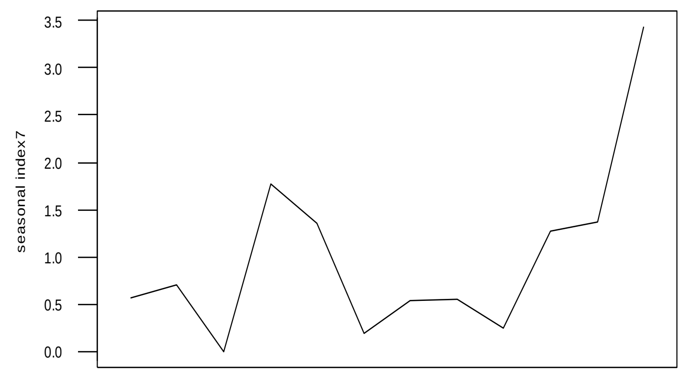

Figure 8: Seasonal Indices Plot of total production yield of finished products (P_7)

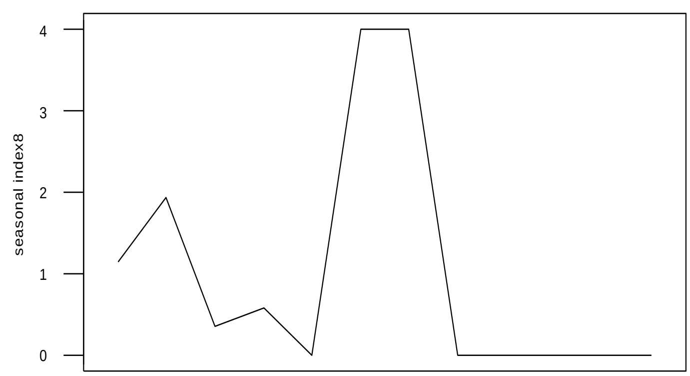

Figure 9: Seasonal Indices Plot of total production yield of finished products (P_8)

Regression Analysis

Using Table 1, where interest is on designing a model that expresses the nature of relationship between the month and product sales. Here month is the dependent variable and p1, p2, p3, p4, p5, p6, p7, p8, are the independent variables. The proposed model is:

$$\text{Month} = \beta_0 + \beta_1\, p1 + \beta_2\, p2 + \beta_3\, p3 + \beta_4\, p4 + \beta_5\, p5 + \beta_6\, p6 + \beta_7\, p7 + \beta_8\, p8 + \varepsilon$$

Regression Analysis

The regression equation is:

$$\text{month} = 41162 - 0.00048\, p1 - 0.00056\, p2 - 0.00240\, p3 - 0.00940\, p4 + 0.0734\, p5 - 0.00163\, p6 + 0.00645\, p7 - 0.159\, p8$$

Predictor	Coef	StDev	T	P
Constant	41161.9	35.8	1151.25	0.000
p1	-0.000484	0.003411	-0.14	0.888
p2	-0.000560	0.002206	-0.25	0.801
p3	-0.002404	0.003812	-0.63	0.534
p4	-0.009402	0.004034	-2.33	0.027
p5	0.07345	0.06712	1.09	0.284
p6	-0.001625	0.005057	-0.32	0.750
p7	0.006452	0.003907	1.65	0.110
p8	-0.15891	0.05991	-2.65	0.013

S = 89.94 R-Sq = 45.2% R-Sq(adj) = 29.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	180389	22549	2.79	0.022
Error	27	218421	8090		
Total	35	398810			

Source	DF	Seq SS
p1	1	51399
p2	1	129
p3	1	4338
p4	1	25982
p5	1	4708
p6	1	32
p7	1	36887
p8	1	56913

Unusual Observations

Obs	p1	month	Fit	StDev Fit	Residual	St Resid
13	7113	40909.0	41091.8	30.3	-182.8	-2.16R
14	7284	40940.0	41143.9	28.6	-203.9	-2.39R
24	18452	41244.0	41206.4	78.0	37.6	0.84 X

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

Regression Analysis

Here product total is the dependent variable and p1, p2, p3, p4, p5, p6, p7, p8, are the independent variables. The proposed model is

Month = $\beta_0 + \beta_1$ p1 + β_2 p2 + β_3 p3 + β_4 p4 + β_5 p5 + β_6 p6 + β_7 p7 + β_8 p8 + ε

The regression equation is

pT = 892 + 0.912 p1 + 1.03 p2 + 1.02 p3 + 0.766 p4 - 2.17 p5 + 1.21 p6
+ 1.16 p7 + 2.36 p8

Predictor	Coef	StDev	T	P
Constant	892	1221	0.73	0.471
p1	0.9119	0.1165	7.83	0.000
p2	1.03403	0.07534	13.73	0.000
p3	1.0177	0.1302	7.82	0.000
p4	0.7658	0.1378	5.56	0.000
p5	-2.169	2.292	-0.95	0.353
p6	1.2133	0.1727	7.03	0.000
p7	1.1570	0.1334	8.67	0.000
p8	2.357	2.046	1.15	0.260

S = 3072 R-Sq = 98.7% R-Sq(adj) = 98.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	19346210216	2418276277	256.28	0.000
Error	27	254770584	9435948		
Total	35	19600980800			

Source	DF	Seq SS
p1	1	15982719459
p2	1	1396196505
p3	1	309447945
p4	1	622462647
p5	1	13011761
p6	1	312002170
p7	1	697852874
p8	1	12516856

Unusual Observations

Obs	p1	pT	Fit	StDev Fit	Residual	St Resid
24	18452	60455	61789	2665	-1334	-0.87 X
30	29097	60997	74575	1615	-13578	-5.20R

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

III. DISCUSSION OF RESULTS

This is based on the results found from the analysis, and also the tables and charts developed:

- One can observe that that in general, there is always a sharp increase of product total in the month of April and a sharp fall in the month of October and November concurrently for over three years.
- It can also observe that the least product total indices are usually in the month of Septembers. It was expressed graphically.
- From the analysis of variance, it can be observed that the p-value= 0.022 when month is dependent and p1-p8 are the independent attributes which is less than the α-value at α=0.05; we will reject the null hypothesis (H_0), hence there is a significant difference. Also, the coefficient of determination for the model obtained is r^2= 45.2% which means that 45.2% is the proportion of variation in the dependent variable (month) that is explained by the variation in the independent variable (product sales). Hence the model is poor for predicting the month of product sales, since the coefficient of determination shows a week relationship.
- From the analysis of variance also, it can be observed that the p-value= 0.000 when product total is dependent and p1-p8 are the independent attributes which is less than the α-value at α=0.05; we will reject the null hypothesis (H_0); hence there is a significant difference. Also, the coefficient

of determination for the model obtained is r^2= 98.7% which means that 98.7% is the proportion of variation in the dependent variable (month) that is explained by the variation in the independent variable (product sales). Hence the model is fit for predicting the product total, since the coefficient of determination shows a strong relationship.

IV. CONCLUSION

In conclusion, a close examination of the production pattern and the behavior of the production system based on the data analyses shows that the production industry is organizing production with a clear focus to meet the customers' requirements and stiff competitors in the plastic manufacturing industry. However, greater percentages of the customers are not served as and when due leading to queues and waiting before customers are served. The tool developed can help the company to remedy this situation.

REFERENCES

[1]. Chatfield, C. (1996). *The Analysis of Time Series: An Introduction* (5th Ed.). London: Chapman & Hall.

[2]. Christensen, Ronald (2002). *Plane Answers to Complex Questions: The Theory of Linear Models* (Third ed.). New York: Springer. ISBN 0-387-95361-2.

[3]. Douglas I. Feiring (1990) Lieutenant Colonel, United States Marine Corps B.A., Virginia Polytechnic Institute and State University, pp. 5 – 8.

[4]. Friston, K.J. Holmes, A.P. Worsley, K.J. Poline, J.-B. Frith C.D. and Frackowiak R.S.J. (1995). "Statistical Parametric Maps in functional imaging: A general linear approach". *Human Brain Mapping* **2**: 189–210. doi:10.1002/hbm.460020402.

[5]. Gujarati, D. N. (2003). *Basic Econometrics* (4th Ed.). New York: McGraw-Hill. Headquarters Marine Corps, Enlisted Assignment Branch, MMEA-5 (2005). "ESGM Overview" Unpublished Briefing.

[6]. Mardia, K.V. Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press. ISBN 0-12-471252-5.

[7]. Rawlings, John O.; Pantula, Sastry G.; Dickey, David A., eds. (1998). *Applied Regression Analysis*. Springer Texts in Statistics. doi:10.1007/b98890. ISBN 0-387-98454-2.

[8]. Wichura, Michael J. (2006). *The coordinate-free approach to linear models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. pp. xiv+199. ISBN 978-0-521-86842-6, ISBN 0-521-86842-4. MR2283455.

[9]. Yaffee, R. A. (2000). *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*. San Diego: Academic Press