Journal of the Nigerian Association of Mathematical Physics
Volume **20 (March, 2012), pp 467 – 476**
© *J. of NAMP*

Estimating Time Series Trend Using Linear Programming Techniques On The Sales Data Of Kaduna Coca-Cola Bottling Company

[1]Aliyu, I. and [2]Sani, B.

[1]Department of Mathematics, Statistics and Computer Science,
Kaduna Polytechnic, Kaduna, Nigeria
[2]Department of Mathematics,
Ahmadu Bello University, Zaria, Nigeria

Abstract

In this paper, we apply linear programming techniques to estimate trend of the sales data of Kaduna Coca-Cola Bottling Company. The trend component of any data has two important properties: smoothness and fidelity (closeness to the data). The linear programming (LP) solution is a monotone sequence which optimizes some weighted combination of both properties.

1.0 Introduction

It is always important to determine the trend of any time-series data which is non-stationary. Trend is defined as a sustained and systematic variation over a long period of time: see Dagum C. and Dagum EG [1]. It is usually represented by some smooth mathematical function such as a polynomial and it is a very important component in determining the forecast of any time series data.

Two important properties of a good trend component estimate are smoothness and fidelity (closeness to the data). Optimizing any of the two measures produces a different trend estimate. If one optimizes smoothness, one deviates from the original data. It is therefore better to estimate the trend based on the trade-off between the two conflicting properties. The first author to consider this trade-off was Whittaker [2] when he was approximating a series Z_i (time-series), i = 1, 2, … N (where N is the number of observations) by a smooth function T_i, i = 1, 2, …. N (Where T_i is the trend during its time).

Whittaker's solution [1] balances the trade-off between smoothness and fidelity by solving the optimization problem given by:

$$Min \left\{\sum_{i=1}^{N}[Z_i - T_i]^2 + \lambda \sum_{i=1}^{N-K}[\Delta^K T_i]^2\right\} \quad \ldots\ldots\ldots\ldots\ldots.. (1)$$

Where K and λ are user – specified constants and Δ is the difference operator. The first term measures the fidelity of the solution to the original data and the second term measures the smoothness of the solution. Whittaker's approach has some inherent problems in that it requires the optimization of a non-linear function which in general may be quite hard. Mosheiov and Raveh [3] introduced a different approach. They used the sum of absolute second order differences as the measure of smoothness and used sum of absolute deviations as their measure of fidelity. Their approach is outline below:

2.0 Estimation Using Linear Programming Approach

Let $Z = (Z_1, Z_2, \ldots., Z_n)$ be the original time-series (Alternatively, let Z be the seasonality adjusted series, i.e. the trend component with irregularities. In this case the series Z can be obtained by any seasonality adjusting procedure. See

[1]Corresponding author: ***Aliyu, I.*** E-mail: -, Tel.: +2348028401256

Raveh [4] for instance). They assumed that Z is generally monotone, i.e. $Z_i \leq Z_{i+1} (Z_i \geq Z_{i+1})$, for most consecutive observations. The aim is to estimate $T = (T_i, \dots, T_N)$ the trend component of Z.

The measure of fidelity is then given as:

$$MF = \sum_{i=1}^{N} |Z_i - T_i| \quad \text{(2)}$$

and for smoothness, they used the sequence of absolute second order differences. Thus

$$|\Delta T_i| = |T_{i+1} - T_i|, i = 1, \dots, N-1,$$

then

$$|\Delta^2 T_i| = |\Delta T_{i+1} - \Delta T_i|,$$
$$= |T_{i+2} - T_{i+1} - (T_{i+1} - T_i)|$$
$$= |T_{i+2} - 2T_{i+1} + T_i|,, i = 1, \dots, N-2,$$

A measure for smoothness (SM) is then given by the sum of the absolute second order differences. i.e.

$$MS = \sum_{i=1}^{N-2} |\Delta^2 T_i| \quad \text{(3)}$$

In order to estimate trend from a time-series data using linear programming techniques, they defined an objective function which is a linear combination of the two factors i.e. measure of fidelity and measure of smoothness.

This is given as:

$$F = \alpha MF + (1-\alpha) MS \quad \text{(4)}$$

for

$$0 \leq \alpha \leq 1$$

Hence the objective function is given by

$$\alpha \sum_{i=1}^{N-2} |Z_i - T_i| + (1-\alpha) \sum_{i=1}^{N-2} |T_{i+2} - 2T_{i+1} + T_i|.$$

This expresses the trade-off between fidelity and smoothness. Monotonicity is however imposed through a set of constraints: within each time segment, the estimated trend T is required to be perfectly montone even if the original data (or the seasonality adjusted data) of that segment of the series is only approximately montone. The estimate trend is therefore, the monotone vector that minimizes equation (4), namely the solution of the following optimization problem.

$$Min\{\alpha \sum_{i=1}^{N-2} |Z_i - T_i| + (1-\alpha) \sum_{i=1}^{N-2} |T_{i+2} - 2T_{i+1} + T_i|\}$$

Subject to

$$T_i \leq T_{i+2}, i = 1, \dots, N$$
$$T_i \geq 0, i = 1, \dots, N \quad \text{(5)}$$

The problem above, is a non-decreasing monotonicity because of the sign of absolute value in the given objective function. Hence it is not a linear function. To convert to a linear one, we can introduce some variables and also change the absolute sign. Since any value in the absolute sign whether positive or negative will become positive, we can define

$$U_i - V_i = Z_i - T_i, \quad U_i \geq 0, V_i \geq 0, i = 1, \dots, N$$
$$X_i - Y_i = T_{i+2} - 2T_{i+1} + T_i, \; X_i \geq 0, Y_i \geq 0, i = 1, \dots, N-2 \quad \dots\dots (6)$$

This idea of eliminating absolute values from objective functions to make them linear was originally presented by Charles et al [5]

After substitution of the new variables i.e. equation (6) into equation (5) above, the following linear program is obtained

$$Min\{\alpha \sum_{i=1}^{N} (U_i + V_i) + (1-\alpha) \sum_{i=1}^{N-2} (X_i + Y_i)\}$$

Subject to

$$U_i - V_i = Z_i - T_i, \; i = 1, \dots, N$$
$$X_i - Y_i = T_{i+2} - 2T_{i+1} + T_i, \; i = 1, \dots, N-2 \quad \text{(7)}$$
$$T_i \leq T_{i+1}, i = 1, \dots, N-1$$
$$U_i, V_i, T_i \geq 0, i = 1, \dots, N$$
$$X_i, Y_i \geq 0, i = 1, \dots, N-2$$

To get the number of variables in (7) as a linear programming problem, we use the formula, $5N - 4$ where N is the number of observations. This is because the number of different variables in the problem is five and each one is counted up to N excluding the X_i, Y_i non-negativity constraints, which are counted up to $N - 2$. That is we get $3N + 2(N - 2) = 5N - 4$. Similarly, to obtain the number of constraints, the formula $3N - 3$ is used due to the fact that one of the constraints goes up to N, another ones goes up to $N - 2$, while the last one goes up to $N - 1$, i.e. we get $N + (N - 1) + (N - 2) = 3N - 3$.

3.0 Application Of Mosheiov And Raveh's Approach To The Data Of Kaduna Bottling Company

We tested the new linear programming (LP) method for trend estimation on data from Kaduna Bottling Company (Coca-Cola) Sales for four years from January 1999 to December 2002 (see data in Table 1).

Table 1: The Originaal Data "Sales Of Nigeria Bottling Company, Kaduna" (January 1999 – 2002)

MONTH	**1999**	**2000**	**2001**	**2002**
JAN	454022	499814	346018	317073
FEB	479432	368002	426803	456009
MAR	609650	406295	712870	726128
APR	456333	377009	546061	517752
MAY	432749	316468	455402	536508
JUN	567722	338972	540607	592768
JUL	377785	309692	415800	431844
AUG	397095	261364	355708	370357
SEP	520383	421815	550744	505699
OCT	421516	360712	396844	437052
NOV	481015	481015	414465	455056
DEC	549014	549014	616264	753239

We initially estimated the trends by using $\alpha = 0$ and $\alpha = 1$ with the help of LINDO Software (we obtain it from internet, i.e., www.lindo.com). After getting the results which represented the trends, we plotted the curves using the original data as well as the trend line. This confirmed to us that the trend pattern matches the data movement. This is because if $\alpha = 1$, then the objective function in the system (5), hence system (7) reduces to

$Min\{\sum_{i=1}^{N} |T_i - Z_i|\}$ which is the best estimate in terms of fidelity.

Similarly if $\alpha = 0$, then the objective function becomes

$Min\{\sum_{i=1}^{N-1} \Delta^2 T_i\}$ which is the best estimate in terms of smoothness.

The graphs we get in the two cases are in figure 1, for $\alpha = 0$ and in figure 2 for $\alpha = 1$.

Putting the above data (i.e. Table 1) into our linear programming model i.e. equation 7, we assigned different $\alpha -$ values as in figure 1, and obtained the trend values.

Our data was large for the computer software because the software was for research purpose and not for commercial use. Hence we had to divide the data into two parts and solve each separately.

Thus, for $\alpha = 0.1$, $1 - \alpha = 0.9$ and $N = 38$ i.e. $i = 1, \ldots, 38$. The number of variables is $5N - 4 = 186$ and the number of constraints is $3N - 3 = 111$.

Hence, our linear program becomes

$$Min\{0.1 \sum_{i=1}^{38} (U_i - V_i) + 0.9 \sum_{i=1}^{36} (X_i + Y_i)\}$$

Subject to

$$U_i - V_i + T_i = Z_i, \quad i = 1, \dots, 38$$
$$X_i - Y_i + T_{i+2} + 2T_{i+1} - T_i = 0, \quad i = 1, \dots, 36$$
$$T_i - T_{i+1} \leq 0, i = 1, \dots, 37$$

$$U_i, V_i, T_i \geq 0, i = 1, \dots, 38$$

$$X_i, Y_i \geq 0, i = 1, \dots, 36$$

Solving the above linear programming problem with the help of LINDO, we estimated the trends. We tested three different values of α, i.e. $\alpha = 0.1$, $\alpha = 0.2$, $\alpha = 0.5$, and $\alpha = 0.9$. The graphs of the data and the estimated trends are in figure 3, 4, 5 and 6 respectively. For more analysis of the trend estimation, refer to Aliyu [6].

From the results obtained, we see that for $\alpha = 0$, the trends estimated are all zeros and as such the trends plotted on the graph are all lined up on the x-axis. For $\alpha = 1$, the trend line is fluctuating due to the results obtained from the estimated trend, we have some zeros in the values which is why in the plotted graph, we see the trend line starts as a straight line and later on goes up and comes down to the x-axis and so on. For $\alpha = 0.9$, the estimated trends are almost similar to those of $\alpha = 1$ and as such the graph plotted with the trend line is almost the same, see figure 2 and 6.

For $\alpha = 0.1$, the values of the estimated trend are almost equal and as such the plotted trend line is almost a horizontal straight line graph, see figure 3. For $\alpha = 0.2$, the plotted graph is similar to that of $\alpha = 0.1$, see figure 4. For $\alpha = 0.5$, the trend estimated is increasing though not very rapidly, see figure 5,

As we can observe from the analysis, for small values of α the trend approximates a straight line, and when α increases, it assumes a step-function form.

Conclusion

The main objective of conducting this study is to highlight how to estimate the trend component of a time series using linear programming approach. We have applied the technique to solve a trend problem of the data obtained from Kaduna Coca-Cola Company Sales of January 1999 – December 2002. Some of the advantages of using the new method include:

1. The trend estimation problem is reduced to a moderate linear program with the number of variables and constraints increasing linearly according to the length of the time series data.
2. The objective function is a tradeoff which balances between smoothness and fidelity so it does not cling solely to one of the properties.
3. The linear programming solution provides a unique definition of the trend component.
4. Because of the recent developments, there are some computer software packages available that can be employed to solve linear programming problems. This method of estimating trend by LP is simpler, more accurate and faster in terms of trend estimate of monotone time series data.
5. With this new procedure, one can solve many problems within a short period of time unlike the other methods such as moving average, freehand method, method of least squares and so on, which are tedious and very slow, in some cases, see Gupta S.P [7].

Table 2: The Estimated Trend Of Table 1 For $\alpha = 0$

	1999	2000	2001	2002
JAN	0	0	0	0
FEB	0	0	0	0
MAR	0	0	0	0
APR	0	0	0	0
MAY	0	0	0	0
JUN	0	0	0	0
JUL	0	0	0	0
AUG	0	0	0	0
SEP	0	0	0	0
OCT	0	0	0	0
NOV	0	0	0	0
DEC	0	0	0	0

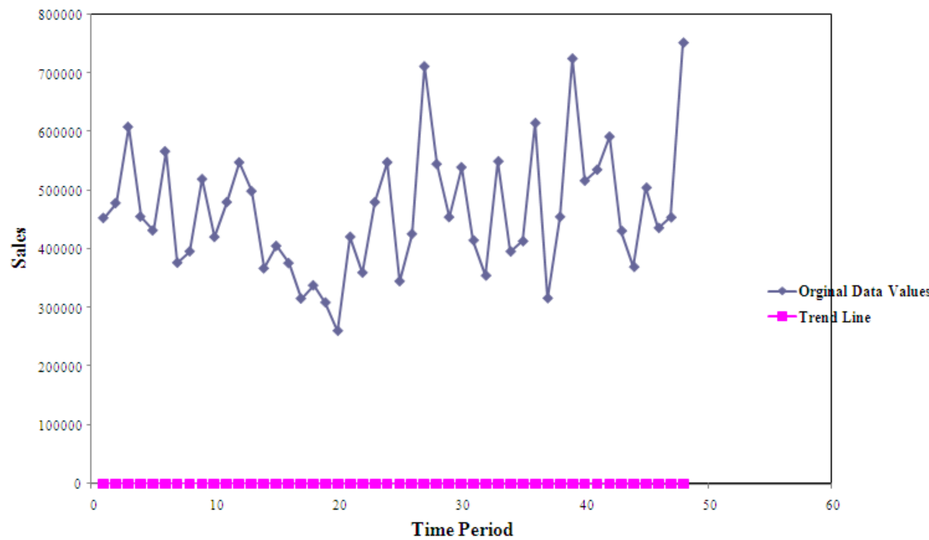

Figure 1: The Original data "Sales of Coca-Cola Company Kaduna Zone and Estimated trend for $\alpha = 0$" (January, 1999 – December 2002).

Table 3: The Estimated Trend Of Table 1 For $\alpha = 1$

	1999	2000	2001	2002
JAN	421576	421516	421516	421516
FEB	421516	421516	421516	421516
MAR	421516	421516	421516	421516
APR	421516	421516	421516	421516
MAY	421516	421516	421516	421516
JUN	421815	421815	481015	481015
JUL	481015	481015	546061	546061
AUG	555402	555402	414465	414465
SEP	414465	414465	414465	414465
OCT	414465	456009	0	0
NOV	0	0	0	0
DEC	0	0	0	0

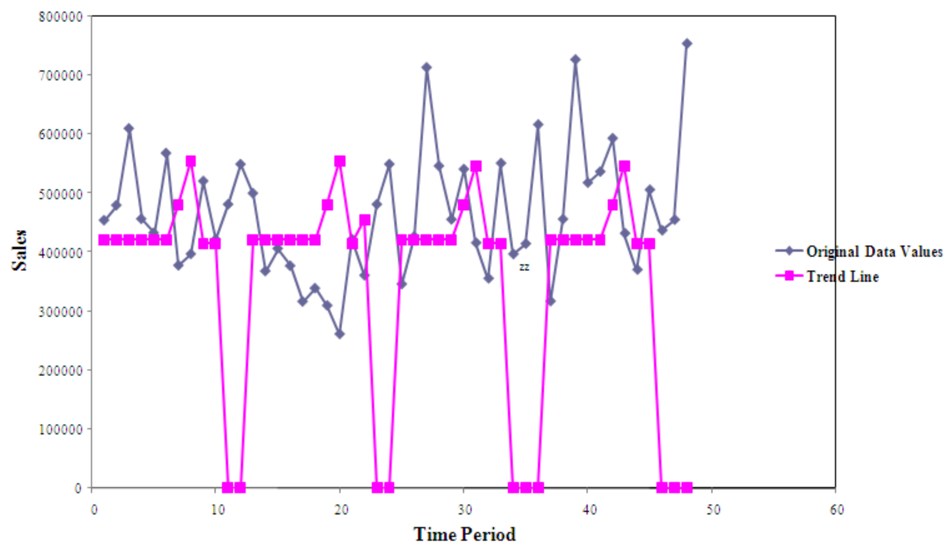

Figure 2: The Original data "Sales of Coca-Cola Company Kaduna Zone and Estimated trend for $\alpha = 1$" (January, 1999 – December 2002).

Table 4: The Estimated Trend Of Table 1 For α = 0.1

	1999	2000	2001	2002
JAN	432749	432749	432749	432749
FEB	432749	432749	432749	432749
MAR	432749	432749	432749	536508
APR	432749	432749	432749	536508
MAY	432749	432749	432749	536508
JUN	432749	432749	432749	536508
JUL	432749	432749	432749	536508
AUG	432749	432749	432749	536508
SEP	432749	432749	432749	536508
OCT	432749	432749	432749	536508
NOV	432749	432749	432749	536508
DEC	432749	432749	432749	536508

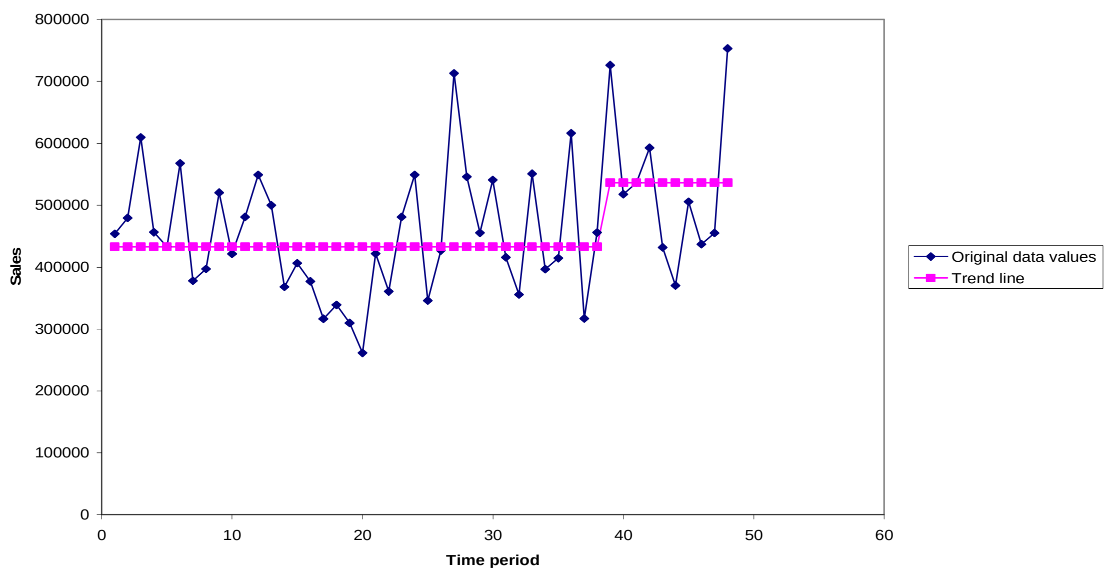

Figure 3: Original data "Sales of Coca-Cola Company Kaduna Zone and estimated trend for $\alpha = 0.1$" (January, 1999 – December 2002).

Table 5: The Estimated Trend Of Table 1 For α = 0.2

	1999	2000	2001	2002
JAN	421815	421815	429860.7	453997.6
FEB	421815	421815	431872.1	456009
MAR	421815	421815	433883.5	505699
APR	421815	421815	435894.9	505699
MAY	421815	421815	437906.3	505699
JUN	421815	421815	439917.7	505699
JUL	421815	421815	441929.1	505699
AUG	421815	421815	443940.5	505699
SEP	421815	421815	445951.9	505699
OCT	421815	423826.4	447963.3	505699
NOV	421815	425837.8	449974.8	505699
DEC	421815	427849.3	451986.2	505699

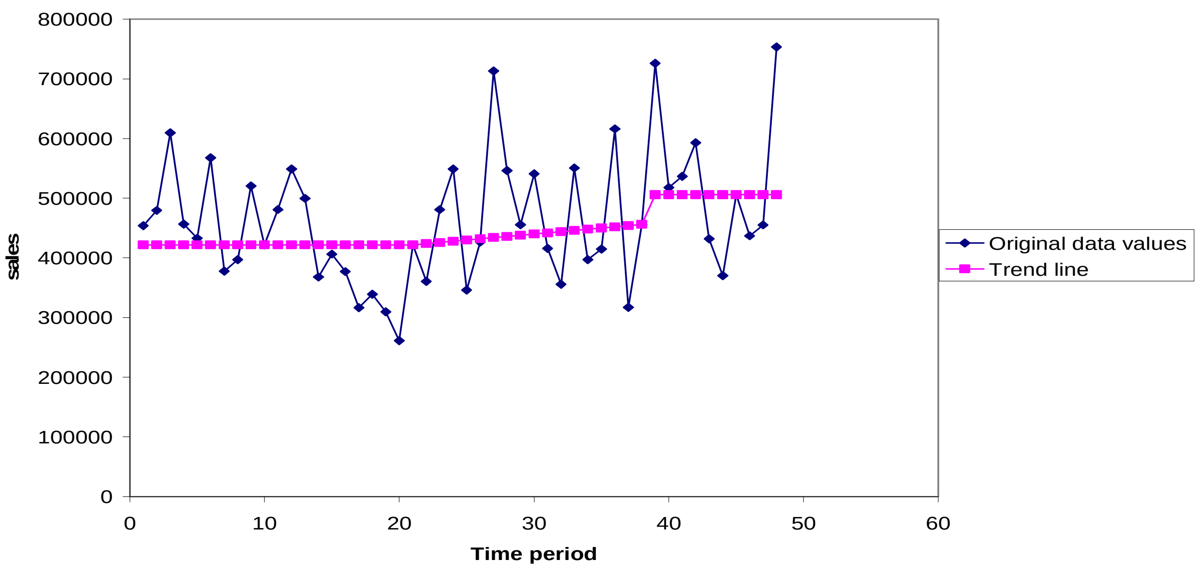

Figure 4: Original data "Sales of Coca-Cola Company Kaduna Zone and estimated trend for $\alpha = 0.2$ (January 1999 – December 2002).

Table 6: The Estimated Trend Of Table 1 For α = 0.5

	1999	2000	2001	2002
JAN	421516	421516	441354.4	456009
FEB	421516	421516	446239.3	456009
MAR	421516	421516	451124.2	456009
APR	421516	421516	456009	456009
MAY	421516	421516	456009	456009
JUN	421516	421516	456009	456009
JUL	421516	421516	456009	456009
AUG	421516	421516	456009	456009
SEP	421516	421815	456009	456009
OCT	421516	426699.8	456009	456009
NOV	421516	431584.7	456009	456009
DEC	421516	436469.6	456009	753239

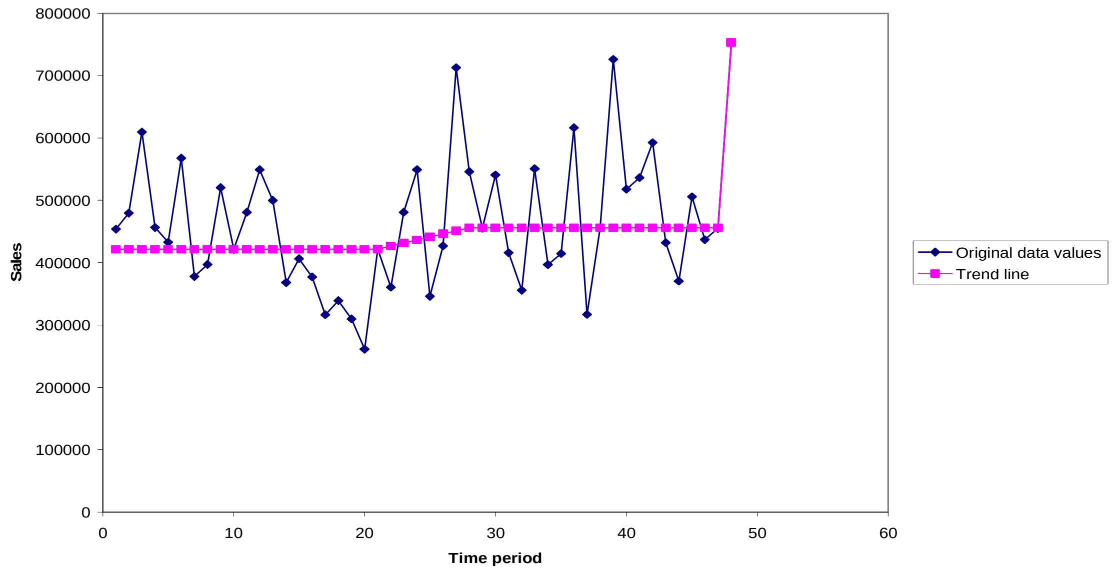

Figure 5: Original data "Sales of Coca-Cola Company Kaduna Zone and estimated trend for $\alpha = 0.5$ (January 1999 – December 2002).

Table 7: The Estimated Trend Of Table 1 For $\alpha = 0.9$

	1999	2000	2001	2002
JAN	421815	421815	421815	421815
FEB	421815	421815	421815	421815
MAR	421815	421815	421815	421815
APR	421815	421815	421815	421815
MAY	421815	421815	421815	421815
JUN	421815	421815	426803	426803
JUL	426803	426803	546061	546061
AUG	546061	546061	414465	414465
SEP	414465	414465	414465	414465
OCT	414465	456009	0	0
NOV	0	0	0	0
DEC	0	0	0	0

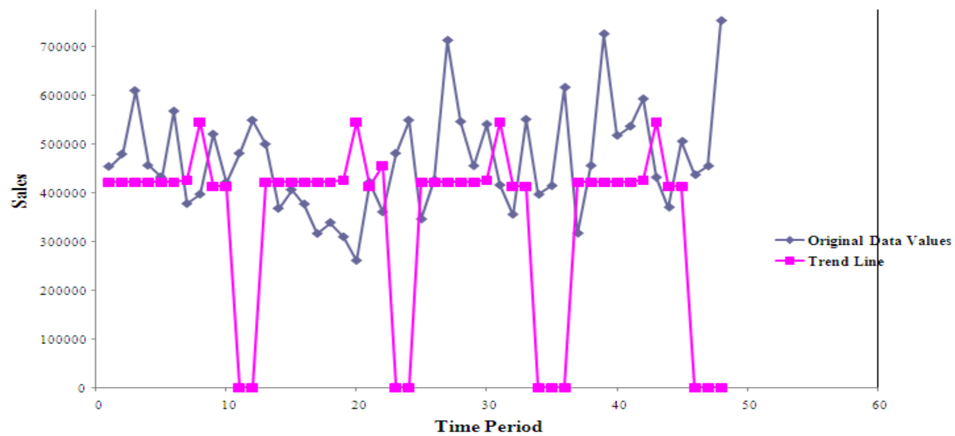

Figure 6: Original data "Sales of Coca-Cola Company Kaduna Zone and estimated trend for $\alpha = 0.9$ (January 1999 – December 2002).

References

[1] Dagum, C. and Dagum, E. G. (1987), "The Entry Trend" in the Encyclopedia of Statistical Science. Vol. 9 (Kotz and Johnson eds.), 321-324. Wiley, New York.

[2] Whittaker, E.T. (1923), On a new Method of Graduation, *Proceedings Edinburgh Mathematics. Society*, 63-75.

[3] Mosheiov, G. and Raveh A (1997), *Journal of the Operational Research Society* 48, 90-96.

[4] Raveh A. (1985), The Chatfield-Prothero Case Study: Another Look, *Journal of Forecasting*, 4: 123-131.

[5] Charles A, Cooper WW and Fergusson R.O. (1955), Optimal Estimation of Executive Compensation by Linear Programming, *Management Science* 1:138-151.

[6] Aliyu I. (2006), Estimating Time Series Trend Using Linear Programming Techniques: A Case Study of Nigeria Bottling Company, Kaduna. An Unpublished M.Sc. Thesis, Ahmadu Bello University Zaria, Nigeria.

[7] Gupta, S.P. (2001): Statistical Method, Sultan Chand and Son, New Delhi, 601-621.