Journal of Physical Science and Innovation **Volume 6, No. 1, 2014**
ISSN: 2277-0119

PREDICTING GOALS IN ENGLISH PREMIER LEAGUE: A CASE STUDY OF MANCHESTER UNITED FOOTBALL CLUB

Ojo, J.F.
Department of Statistics,
University of Ibadan, Ibadan, Nigeria.
E-mail: jfunminiyiojo@yahoo.co.uk

ABSTRACT

This study develops deterministic and probability models particularly additive and bilinear models with a view to determine the future goals in Manchester United Football Club. The nonlinear and non stationary nature of the series (goals) calls for the use of non linear time series model in addition to the deterministic model. The parameters of these models were estimated using the appropriate technique. The appropriate technique also was used to select the best model which was used to determine the future goals in Manchester United Football Club. For the deterministic model the future goals were determined and for the probability model the direction of goals in future was also determined. The future goals from these models will guide the policy maker (in football match) in decision taken.

Keywords: **Football Match, Policy Makers, Additive and Bilinear Models, Manchester United**

INTRODUCTION

Manchester United Football Club is an English professional football club, based in Old Trafford, Greater Manchester, which plays in the Premier League. Founded as Newton Heath LYR Football Club in 1878, the club changed its name to Manchester United in 1902 and moved to Old Trafford in 1910. Manchester United has won twenty League titles, eleven FA Cups, four League Cups, and twenty FA Community Shields. The club has also won three European Cups, one UEFA Cup Winners' Cup, one UEFA Super Cup, one Intercontinental Cup, and one FIFA Club World Cup. In 1998–99, the club won a continental treble of the Premier League, the FA Cup and the UEFA Champions League, an unprecedented feat for an English club. The 1958 Munich air disaster claimed the lives of eight players. In 1968, under the management of Matt Busby, Manchester United was the first English football club to win the European Cup. Alex Ferguson won 28 major honours, and 38 in total, from November 1986 to May 2013, when he announced his retirement after 26 years at the club. Fellow Scot David Moyes was appointed as his replacement on 9

May 2013. Manchester United is the third-richest football club in the world for 2011–12 in terms of revenue, with annual revenue of €395.9 million, and the second most valuable sports team in 2013, valued at $3.165 billion. It is one of the most widely supported football teams in the world. After being floated on the London Stock Exchange in 1991, the club was purchased by Malcolm Glazer in May 2005 in a deal valuing the club at almost £800 million. In August 2012, Manchester United made an initial public offering on the New York Stock Exchange (en.m.wikipedia.org/wiki/Manchester_United_F.C.). In view of the above achievement it is imperative to provide best models and future of goals by this football club.

METHODOLOGY

The theory of stochastic processes plays an important role in the investigation of random phenomena depending on time, a time series is a special kind of stochastic process indexed by time (t), however, other dimensions like space, volume etc are sometimes used to index the series. A time series is a realization of an ensemble, which is the totality of time series generated by the particular underlying process. When the probabilistic structure of all the realizations in an ensemble is the same, we say that the process underlying the series is ergodic. It is these properties of ergodicity and stationary that warrants us to estimate ensemble parameters by the corresponding time averages of just one realization Shangodoyin and Ojo (2002). Time series is a set of statistical observations made sequentially in time, it is possible to observe a time variable at any instant, and thus the temporal intervals between successive members of the series need not be the same. But practice and theory alike, we require the observations to occur at regular intervals. A time series could be considered as a mixture of four components such as trend, seasonal variation, cyclical and irregular variation. In the light of these components of a time series $\{X_t\}$, we can represent a series in two convenient models as some functions of the component. The standard forms are:

(i) $X_t = T_t.S_t.C_t.I_t$ called the multiplicative model and

(ii) $X_t = T_t + S_t + C_t + I_t$ called the additive model Shangodoyin and Ojo (2002)

The series (goals) of Manchester United exhibit additive measure, as a result we shall employ additive model in this study.

Linear Curve

It is used to estimate the trend curve of a set of observations measured over time, when the trend exhibits a straight line pattern. Consider the simple linear curve given as:

$X_t = a + bt + e_t$.To fit this curve to a set of data, when t is the time period, the least squares method enables us to estimate the parameters a and b. This is done by minimizing error sum of squares

$$Q = \sum e_t^2 = (X_t - a - bt)^2$$

With respect to a and b. That is

$$\frac{dQ}{da} = 0 \text{ and } \frac{dQ}{db} = 0,$$

Gives the set of normal equations as follows:

$$\sum_{t=1}^{n} X_t = nb_0 + b_1 \sum_{t=1}^{n} t$$

$$\sum_{t=1}^{n} tX_t = b_0 \sum_{t=1}^{n} t + b_1 \sum_{t=1}^{n} t^2$$

Simplifying the normal equations to give $\hat{a} = \overline{X} - b\bar{t}$,

$$\hat{b} = \frac{n\sum tX_t - \sum t \sum X_t}{n\sum t^2 - (\sum t)^2} \text{ or } \hat{b} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(t - \bar{t})}{\sum_{t=1}^{n}(t - \bar{t})^2}$$

Where $\overline{X} = \frac{\sum X_t}{n}$ and $\bar{t} = \frac{\sum t}{n}$. Hence, the fitted trend is $T_t = X_t = \hat{a} + \hat{b}t$ Ojo (2010).

Suppose a series X_t has an additive relationship with the time series components, that is, $X_t = T_t + S_t + C_t + I_t$. Then the seasonal variation at time t is $S_t = X_t - T_t$. From this relationship seasonal index is determined. Very often we have series of measurements which are affected by some time series components, such as trend and seasonal variation; these are adjusted for in forecasting. Any change in the more sensitive series will anticipate the corresponding change in the components therein, and can be used as a forecasting indicator. Generally, the forecasting for period p is $X_p = T_p + SI_p$ for the additive model where SI_p is the seasonal index for period p Shangodoyin and Ojo (2002).

One-dimensional Autoregressive Integrated Moving Average Bilinear Time Series Models

We define one-dimensional autoregressive integrated moving average bilinear time series models as follows:

$\psi(B)X_t = \phi(B)\nabla^d X_t + \theta(B)e_t + \left(\sum_{k=1}^{r} b_{k1}X_{t-k}\right)e_{t-1}$, denoted as BL (p, d, q, r, 1),

Where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 .... - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 .... - \theta_q B^q$, and

$$X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} + e_t - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} + \left(\sum_{k=1}^{r} b_{k1}X_{t-r}\right)e_{t-1}$$

$\phi_1,...,\phi_p$ are the parameters of the autoregressive component; $\theta_1,...\theta_q$ are the parameters of the associated error process; $b_{11},........., b_{r1}$ are the parameters of the non-linear component and $\theta(B)$ is the moving average operator. p is the order of the autoregressive component; q is the order of the moving average process; *r1* is the order of the nonlinear component and $\psi(B) = \nabla^d \phi(B)$ is the generalized autoregressive operator. ∇^d is the differencing operator and d is the degree of consecutive differencing required to achieve stationary. Also e_t are independently and identically distributed as $N(0, \sigma_e^2)$ Ojo (2010)

Algorithm for Fitting One-dimensional Autoregressive Integrated Moving Average Bilinear Time Series Models

For the sake of simplicity, we will break the algorithm down into the following steps:

Step 1

Fit various order of autoregressive integrated moving average (ARIMA) model of the form $X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} + e_t$ where the parameters estimated form the initial values in the estimation of one dimensional autoregressive integrated moving average bilinear time series models.

Step 2

Choose the model for which Akaike Information Criterion (AIC) Akaike (1974) is minimum among various order fitted in step 1.

Step 3

Fit various order of one dimensional autoregressive integrated moving average bilinear model of the form $X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} + b_{11} X_{t-1} e_{t-1} + ... + b_{r1} X_{t-r} e_{t-1} + e_t$ using non linear least square method and choose the model for which AIC is minimum.

Step 4

The model with the minimum AIC is the best one dimensional autoregressive integrated moving average bilinear model.

Predictive Model

The forecasts for the probability model are obtained as follows: Suppose $\{X_t\}$ is a discrete time series and we wish to predict X_{t_0+h} given the semi-infinite realization $(X_{s,} s \leq t_0)$. Let the predictor be $\tilde{X}_{t_0}(h)$. Then it is well known that $E\left[X_{t_0+h} - \tilde{X}_{t_0}(h)\right]^2$ is minimum if and only if $\tilde{X}_{t_0}(h) = E(X_{t_0+h} / X_{s,} s \leq t_0)$. The evaluation of $\tilde{X}_{t_0}(h)$ from the model depends on the unknown parameters. Typically, we substitute the least squares estimates of these parameters, and then calculate the predictors. The predictors thus obtained are denoted by $\tilde{X}_{t_0}(h)$, (h = 1, 2, …) and the error by $\hat{e}_{t_0}(h) = X_{t_0+h} - \hat{X}_{t_0}(h)$ and the mean sum of squares of the errors of the predictors for the period (t_o+h, t_o+ h + 1, …, t_o + h + M) is $\hat{\sigma}_{\hat{e}}^2(h) = \frac{1}{M}\sum_{j=1}^{M} \hat{e}_{t_0+j}^2(h)$ (Ojo 2009, 2012).

RESULTS AND DISCUSSION

The monthly data used was extracted from www.soccerway.com. The study covers a range of twenty one years from 1992 to 2012. The time plot is given below

Time Plot

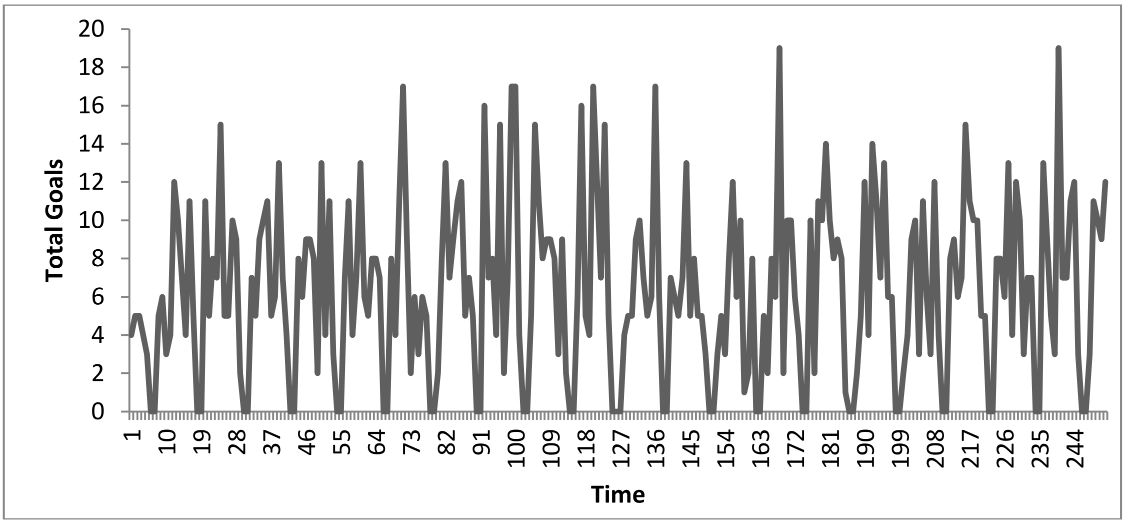

Figure 1: Time Plot Showing Total Goals Scored by Manchester United from 1992 to 2012

Source: Soccerway.com

Best Fitted Models

$\hat{X}_t = 5.898 + 0.003t,\ S.V = X_t - \hat{X}_t$

$\hat{X}_t = 0.132925X_{t-1} - 0.115677X_{t-2} - 0.019656X_{t-3} - 0.018766X_{t-4} - 0.249863X_{t-5} - 0.175653X_{t-6} - 0.175364X_{t-7} + 0.991371e_{t-1} - 0.046419X_{t-1}e_{t-1} - 0.020777X_{t-2}e_{t-1}$

The above models are the trend, seasonal variation and bilinear models respectively. The derived statistics from the bilinear model are given in table1 below.

Table 1: Derived Statistics

	ODARIMABL
Adjusted R^2	0.550753
AIC	5.499894
BIC	5.527905

When fitting various orders of bilinear time series models, the Akaike and Bayesian Information criteria were used to determine the best order. The order at which we had the minimum AIC and BIC were recorded in the table above. The corresponding adjusted r-squared that measure the goodness of fit is also given in the above table.

Future Goals

Table 2: Future Goals for Manchester United Football Club from January 2013 to December 2015 in English Premiership (League)

	2013	2014	2015	Seasonal Index
Jan.	8	8	8	0.822
Feb.	8	8	8	1.390
Mar.	8	8	8	1.197
Apr.	9	9	9	1.956
MAY	5	5	5	-2.190
June	0	0	0	-6.336
July	0	0	0	-6.339
Aug.	7	7	7	0.325
Sep.	7	7	7	0.750
Oct.	8	8	8	1.319
Nov.	8	8	8	1.411
Dec.	12	12	12	5.694

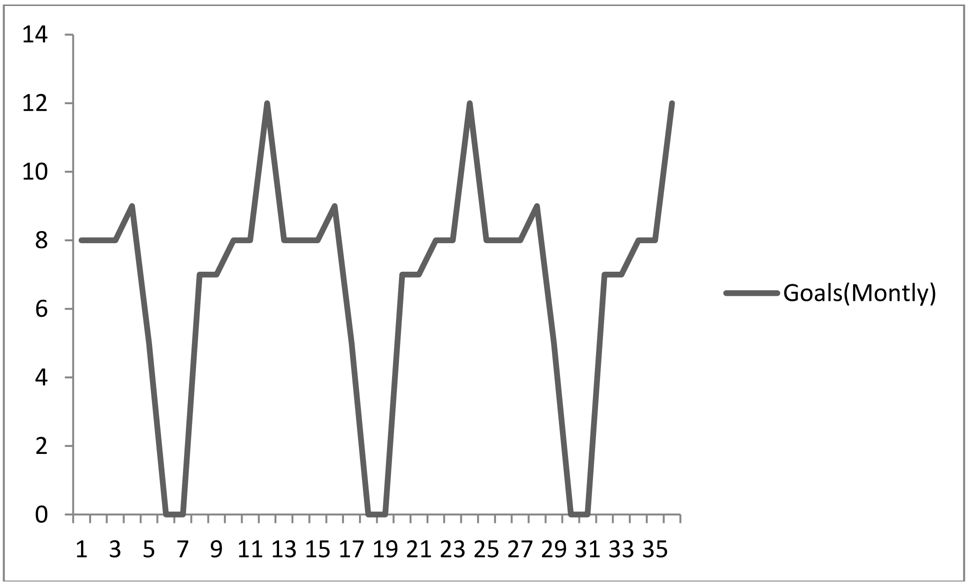

Figure 2: Future Goals Using Deterministic Model for Manchester United Football Club from January 2013 to December 2015

Figure 2 shows that in January 2013, Manchester United will score a total of 8 goals and in December 2015, they will score a total of 12 goals in English Premiership League (EPL). From the original series, which we represented by the time plot in figure 1 above, the month of June and July do not have goals since the club is always on vacation during this period so as to prepare for the next season. Likewise from the future goals represented in figure 2, using the deterministic model the same pattern was retained and we could see that there were no goals in the month of June and July of every year for the club to prepare for the next season. Being a deterministic model the variation in the future goals were constant from year to year. In figure 3 below, the direction of future goals of Manchester United Football Club is shown. From the figure, we have upward movement in the future goals using bilinear model. What is being suggested by this upward movement is that, in the future years there will be increment in the goals scored by this team. With this information, the executive of this team can plan ahead of time and take some decision before this future time.

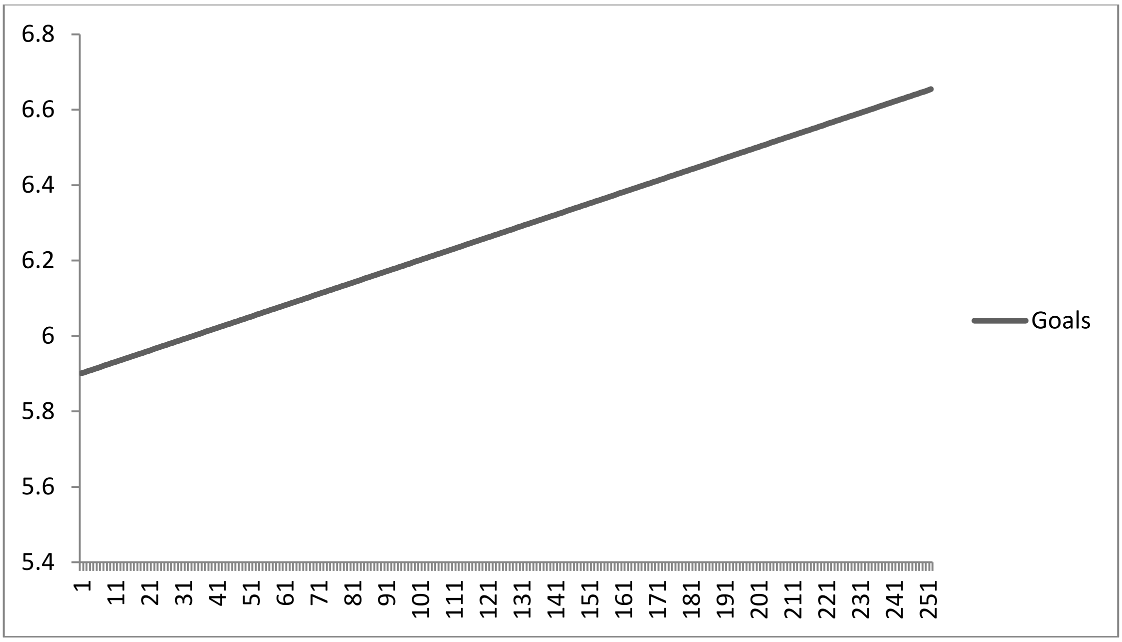

Figure 3: Direction of Future Goals Using Probability (Bilinear) Models

CONCLUSION

In this study we have looked at the goals scored by Manchester United Foot Club and fitted appropriate deterministic and probability models to the data. This was done so as to determine the future goals that would be scored by this club for appropriate action to be taken by the executive of this team now. The nonlinear and non stationary nature of the series (goals) calls for the use of non linear time series model in addition to the deterministic model. The parameters of these models were estimated using non linear least square. Akaike and Bayesian Information Criteria were used to select the best model which was used to determine the future goals in Manchester United Football Club. For the deterministic model the future goals were determined and for the probability model the direction of goals in future was also determined. The future goals from these models will guide the policy maker (in football match) in decision taken.

REFERENCES

Akaike, H. (1974). "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes". *Ann. Inst. Stat. Math.* 26, 363-383.

Ojo J. F. (2009). "The Estimation and Prediction of Autoregressive Moving Average Bilinear Time Series Models with Applications" *Global Journal of Mathematics and Statistics,* Vol. 1 No. 2, 111-117.

Ojo, J.F. (2010). On the Estimation and Performance of One Dimensional Autoregressive Integrated Moving Average Bilinear Time Series Models. *Asian Journal of Mathematics and Statistics* 3(4), 225-236.

Ojo, J.F. (2010). Time and Trend Plots: Application to HIV/AIDS Data. *Journal of Science and Technology Research,* Volume 9, Number 4, Pages 56-62.

Ojo, J.F. (2012). Forecasting in One-Dimensional and Generalized Integrated Autoregressive Bilinear Time Series Models. *Global Journal of Mathematical Sciences.* Volume 11, No 1 & 2, Pages 27 – 34.

Shangodoyin, D.K and Ojo, J.F (2002). "Elements of Time Series Analysis" Rashmed Publication, Ibadan. ISBN 978-8024-32-7.

en.m.wikipedia.org/wiki/Manchester_United_F.C.
www.soccerway.com

Reference to this paper should be made as follows: Ojo, J.F. (2014), Predicting Goals in English Premier League: A Case Study of Manchester United Football Club. *J. of Physical Science and Innovation,* Vol. 6, No. 1, Pp. 54 – 61.