International Journal of Mathematics and Statistics Studies

Vol.5, No.1, pp.45-60, February 2017

Published by European Centre for Research Training and Development UK (www.eajournals.org)

A COMPARATIVE TIME SERIES ANALYSIS OF POINTS SCORED BY ARSENAL FOOTBALL CLUB IN PREMIERSHIP COMPETITIONS USING DESCRIPTIVE AND PROBABILITY MODELING (STOCHASTIC) APPROACHES

[1]Otuonye, Eric L, [2]Okororie Chigozie and [3]Ibeh, Gabriel C.

[1]Department of Statistics, Abia state University Uturu, Nigeria
[2]Registry Department, Alvan Ikoku University of Education, Owerri Nigeria
[3]Mathematics/Statistics Department, Federal Polytechnic, Nekede, Owerri, Nigeria

ABSTRACT : *This paper discusses the performance of Arsenal Football club London from 1893/1894 to 2013/2014 seasons by considering the points scored every season. The method of analysis adopted includes comparing the forecasting strength of two time series analysis approach: Descriptive and Probability modelling approaches. Summary statistics which includes three accuracy measures (Mean Absolute Deviation (MAD), Mean Square Deviation (MSD) and Mean Absolute Percentage Error (MAPE)) of the residuals and forecast errors from the fitted quadratic models and Autoregressive-Integrated Moving Average (ARIMA) (2, 2, 0) models were used for the comparison. The data was assessed for variance stability and it was found that there is need for square root transformation. The forecasting strengths of the two approaches were compared using the forecast errors of the two models. The results of the analyses showed that the descriptive model performed better than the fitted probability model*.

KEYWORDS: Data transformation, Summary statistics, Buys-Ballot table, group averages and standard deviation and Differencing.

INTRODUCTION

Arsenal Football Club is a professional football club based in Holloway London. The Club play in the top flight of English football called the Premier League. It is one of the most successful clubs in English football. They have won so many laurels including 13 first division and premier league titles and 12 FA cups. The club was founded in 1886 and became the first from the south of England to join the football league in 1893. It is also the first London club to reach the Union of European Football Association (UEFA) Champions League final. The club moved to their present Emirates stadium near Holloway in 2006 from its former stadium in Highbury. Soar, et al[1] has it that Arsenal football club was formed as Dial Square in 1886 by workers at the Royal Arsenal in Woolwich, south east London and were renamed Royal Arsenal Shortly afterwards. The club according to Forbes[2] was named the seventh most valuable association football club valued at $1.3 billion in 2015. Arsenal enjoyed periods of successes in the pre second world war II under notable Managers such as Herbert Chapman. According to Sour and Tyler[3] Arsenal appointed Herbert Chapman as Manager in 1925 and his appointment brought their first period of major success by his revolutionary tactics and training and consequently laid the foundation of the club's domination of English football in the 1930's.

The club's success continued even after World War II especially with the appointment of its present manager in 1996. Hughes I[4] has it that the clubs success in the late 1990's and the first

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.5, No.1, pp.45-60, February 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

decade of the 21st century over a great deal owe it to the appointment of Arsene Wenger as manager in 1996. Arsenal has won a second league and cup double in 1997/98 and a third in 2001/2002. In addition, the club reached the final of the 1999/2000 UEFA cup (losing on penalties to Galatasary); were victorious in the 2003 and 2005 FA cups and won the premier league in 2003/2004 without losing a single match, an achievement which earned the side the nickname "The Invincibles". Frazer A [5] said that the club's feat in the 2003/2004 season came within a run of 49 league matches unbeaten from 7th May, 2003 to 24th October 2004, a national record.

A lot of attention has been paid in recent times in the prediction of football matches and that is why Anderson C and Sally D[6] said that the prediction of outcomes in individual football matches is an intricate and mind burgling issue due to the number of goals scored. One of such predictions of soccer results was done by Joseph et al[7] They used the Bayesian Nets to predict the results of Tottenham Hotspur over the period of 1995-1997. The model relied on trends from a specific time period and cannot be extended to later seasons. This limitation notwithstanding, it provided useful background on comparison of models feature selection. Another attempt was done by Rue et al[8]. They used a Bayesian Linear model to predict soccer results by using a time dependent model that investigated the relative strength of attack and defense of each team. In another development, in the modeling of English Premier League, Addae et al[9] modeled the trend of Manchester United Football club in the 1960-2013 English Premiership using ARIMA model and came up with a forecast that the club would lose six (6) games for 2013/2014 season.

In a recent study on forecasting in sports, Yiannakis et al[10] used the method of multivariate ARIMA to predict the outcome of English premier league soccer with a success rate of nine out of 10 for winning, eight out 10 for losing and nine out of 10 for drawing. They hold that a mix of both shared and new variables in different sets of interactions help predict results. The accuracy of the final league tables of domestic football leagues was investigated by Haaren and Davis [11] shows that points can be predicted both before the start of the season and during the course of the season. They performed an empirical evaluation that compares two flavours of the well-established Elo-rating and the recently introduced Pi-rating. They validated the different approaches using a large volume of historical match results from several European football leagues.
The specific objectives of the work are to;

(a) Evaluate the study data and, where necessary, determine the approximating transformationto stabilize the variance.
(b) Fit the appropriate Descriptive and ARIMA model(s) to the transformed series and
(c) Compare the summary statistics based on residuals from
    (i) The fitted models (descriptive and ARIMA) and
    (ii) The forecast errors
(d) To make recommendations based on the results of the analysis)

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

2: Methodology

In this study, secondary data on Arsenal Football Club, London yearly football season points from number of matches played were obtained from official Arsenal Football Club official website (www.arsenalfc.co.uk). The data covered between 1893/1894 football season to 2014/2015 seasons as shown in Table 1.

Table 1: Table of Points scored and matches played per season (by Arsenal Football Club)

Season	Pld	Pts	Season	Pld	Pts	Season	Pld	Pts	Season	Pld	Pts	Season	Pld	Pts	Season	Pld	Pts
2014/2015	38	75	1977/1978	42	52	1933/1934	42	59	1995/1996	38	63	1958/1959	42	50	1910/1911	38	38
2013/2014	38	79	1976/1977	42	43	1932/1933	42	58	1994/1995	42	51	1957/1958	42	39	1909/1910	38	31
2012/2013	38	73	1975/1976	42	36	1931/1932	42	54	1993/1994	42	71	1956/1957	42	50	1908/1909	38	38
2011/2012	38	70	1974/1975	42	37	1930/1931	42	66	1992/1993	42	56	1955/1956	42	46	1907/1908	38	36
2010/2011	38	68	1973/1974	42	42	1929/1930	42	39	1991/1992	42	72	1954/1955	42	43	1906/1907	38	44
2009/2010	38	75	1972/1973	42	57	1928/1929	42	45	1990/1991	38	**831**	1953/1954	42	43	1905/1906	38	37
2008/2009	38	72	1971/1972	42	52	1927/1928	42	41	1989/1990	38	62	1952/1953	42	54	1904/1905	34	33
2007/2008	38	83	1970/1971	42	65	1926/1927	42	43	1988/1989	38	76	1951/1952	42	53	1903/1904	34	49
2006/2007	38	68	1969/1970	42	42	1925/1926	42	52	1987/1988	40	66	1950/1951	42	47	1902/1903	34	48
2005/2006	38	67	1968/1969	42	56	1924/1925	42	33	1986/1987	42	70	1949/1950	42	49	1901/1902	34	42
2004/2005	38	83	1967/1968	42	44	1923/1924	42	33	1985/1986	42	69	1948/1949	42	49	1900/1901	34	36
2003/2004	38	90	1966/1967	42	46	1922/1923	42	42	1984/1985	42	66	1947/1948	42	59	1899/1900	34	36

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

2002/2003	38	78	1965/1966	42	37	1921/1922	42	37	1983/1984	42	63	1946/1947	42	41	1898/1899	34	41
2001/2002	38	87	1964/1965	42	41	1920/1921	42	44	1982/1983	42	58	1938/1939	42	47	1897/1898	30	37
2000/2001	38	70	1963/1964	42	45	1919/1920	42	42	1981/1982	42	71	1937/1938	42	52	1896/1897	30	30
1999/2000	38	73	1962/1963	42	46	1914/1915	38	43	1980/1981	42	53	1936/1937	42	52	1895/1896	30	32
1998/1999	38	78	1961/1962	42	43	1913/1914	38	49	1979/1980	42	52	1935/1936	42	45	1894/1895	30	34
1997/1998	38	78	1960/1961	42	41	1912/1913	38	18	1978/1979	42	48	1934/1935	42	58	1893/1894	28	28
1996/1997	38	68	1959/1960	42	39	1911/1912	38	38									

Key: Pld = Matches Played; Pts = Points gained.
(Arrange from the earliest to the latest)

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.5, No.1, pp.45-60, February 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

It should be noted that seasons 2013/2014 and 2014/2015 were omitted from the analysis to enable us compare forecast and actual results. The data was first divided with number of matches played as there were seasons when the competing teams are not twenty (20) as it now evaluated for need for transformation following the procedure of Akpanta and Iwueze [12] on applying the Bartlett[13] transformation to non-seasonal time series data. The data was assessed for need for transformation. The transformed data was subjected to Descriptive analysis and Box-Jenkins ARIMA modeling procedures to determine the appropriate model. Details of ARIMA modeling procedures can be found in Box and Jenkins [14] and Wei[15].

Finally, the performances of the models are compared using what is referred to in the literature as "Summary statistics" based on the residuals from (i) the fitted models and (ii) the forecast errors. The Summary statistics from the fitted models and forecast error includes; The Mean Square Deviation (MSD), Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE). Their respective mathematical representations are;

$$MSD = \frac{\sum_{i=1}^{m} e_i^2}{m} \qquad (1)$$

$$MAD = \frac{\sum_{i=1}^{m} |e_i|}{m} \qquad (2)$$

$$MAPE = \left(\frac{1}{m} \sum_{i=1}^{m} \left| \frac{e_i}{Z_i} \right| \right) * 100 \qquad (3)$$

where e_i is the residual from the transformed series, m is the number of observations used in the computation and Z is the number of errors obtained after .

Table 2: Buys –Ballot Table of the Original Data/Points Scored by Arsenal FC in a Season

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.5, No.1, pp.45-60, February 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

RESULTS

Period	Block									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	28	34	32	30	37	31	36	36	42	48
2	49	33	37	44	36	38	31	38	38	18
3	49	43	42	44	37	42	33	33	52	43
4	41	45	39	66	54	58	59	58	45	52
5	52	47	41	59	49	49	47	53	54	43
6	43	46	50	39	50	39	41	43	46	45
7	41	37	46	44	56	42	65	52	57	42
8	37	36	43	52	48	52	53	71	58	63
9	66	69	70	66	76	62	83	72	56	71
10	51	63	68	78	78	73	70	87	78	90
11	83	67	68	83	72	75	68	70	73	

	Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**	**Block 6**	**Block 7**	**Block 8**	**Block 9**	**Block 10**	$\overline{X}_i$	$\hat{\sigma}_i$
Periods 1	1.00000	1.13333	1.06667	1.00000	1.23333	0.91176	1.05882	1.05882	1.23529	1.41176	1.11098	0.146484
Periods 2	1.44118	0.97059	0.97368	1.15789	0.94737	1.00000	0.81579	1.00000	1.00000	0.47368	0.97802	0.242673
Periods 3	1.28947	1.13158	1.00000	1.04762	0.88095	1.00000	0.78571	0.78571	1.23810	1.02381	1.01830	0.171042
Periods 4	0.97619	1.07143	0.92857	1.57143	1.28571	1.38095	1.40476	1.38095	1.07143	1.23810	1.23095	0.211191
Periods 5	1.23810	1.11905	0.97619	1.40476	1.16667	1.16667	1.11905	1.26190	1.28571	1.02381	1.17619	0.127082
Periods 6	1.02381	1.09524	1.19048	0.92857	1.19048	0.92857	0.97619	1.02381	1.09524	1.07143	1.05238	0.094443
Periods 7	0.97619	0.88095	1.09524	1.04762	1.33333	1.00000	1.54762	1.23810	1.35714	1.00000	1.14762	0.211713
Periods 8	0.88095	0.85714	1.02381	1.23810	1.14286	1.23810	1.26190	1.69048	1.38095	1.50000	1.22143	0.262279
Periods 9	1.57143	1.64286	1.66667	1.65000	2.00000	1.63158	2.18421	1.71429	1.33333	1.69048	1.70849	0.232515
Periods 10	1.21429	1.65789	1.78947	2.05263	2.05263	1.92105	1.84211	2.28947	2.05263	2.36842	1.92406	0.329988
Periods 11	2.18421	1.76316	1.78947	2.18421	1.89474	1.97368	1.78947	1.84211	1.92105		1.94498	0.151381

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.5, No.1, pp.45-60, February 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

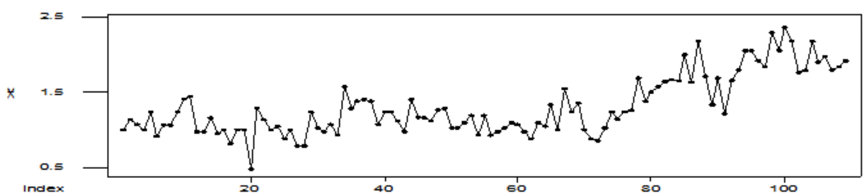


Fig.1: Time plot of the original series

Table 4: Periodic means and standard deviations and their natural logarithms

$\overline{X}_i$	$\hat{\sigma}_i$	$\text{Log}_e(\overline{X}_i)$	$\text{Log}_e(\hat{\sigma}_i)$
1.11098	0.146484	0.105241	-1.92084
0.97802	0.242673	-0.022227	-1.41604
1.01830	0.171042	0.018130	-1.76585
1.23095	0.211191	0.207788	-1.55499
1.17619	0.127082	0.162281	-2.06292
1.05238	0.094443	0.051056	-2.35976
1.14762	0.211713	0.137689	-1.55252
1.22143	0.262279	0.200021	-1.33835
1.70849	0.232515	0.535607	-1.45880
1.92406	0.329988	0.654437	-1.10870
1.94498	0.151381	0.665249	-1.88796

From Table 4, the annual standard deviation ranged from about 0.146484 to about 0.329988, indicating that the variance is not appreciably constant and requires transformation. [15]

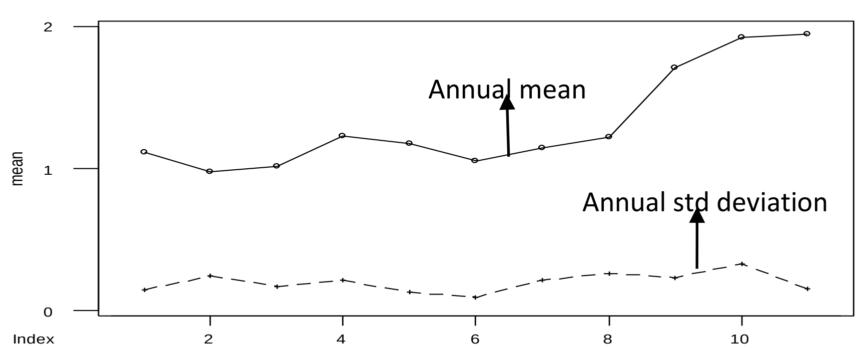


Fig. 2: Plots of the annual means and annual standard deviations

Figure 2, it is evident a change (increase or decrease) in the mean, does not have the same effect in the periodic standard deviation. Consequently, additive model was used in decomposing the series since from Figure 2, the overlaid plot of annual means and annual standard deviation shows that the plots are negatively correlated in the sense that change in mean does not produce the same effect on the standard deviation.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.5, No.1, pp.45-60, February 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

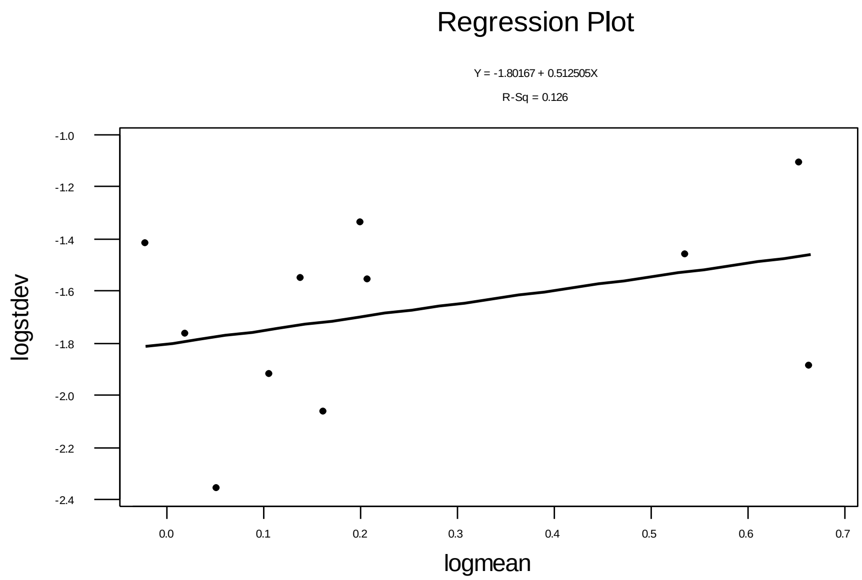


Fig. 3: Regression plot

Furthermore, the slope of the linear regression of natural logarithm of the periodic standard deviation ($\log_e \hat{\sigma}_x$) on the natural logarithm of the periodic means ($\log_e \overline{X}_{i.}$) of the study data was found to be $\hat{\beta}_x = 0.513$ with the standard error 0.4492 and coefficient of determination $R^2 = 0.126$. The p-value associated with it (0.028) indicates that the observed value of $\hat{\beta}$ is significantly different from zero at $\alpha = 0.05$ level of significance, thus confirming that the data requires transformation. Bartlett's transformation for β and summarized by Akpanta and Iwueze[12] show that $\beta = 0.5$ corresponds to square root transformation. Hence the observed series was subjected to square root transformation. Time plot of the square root transformed data (Y_t) is shown in Figure 4

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

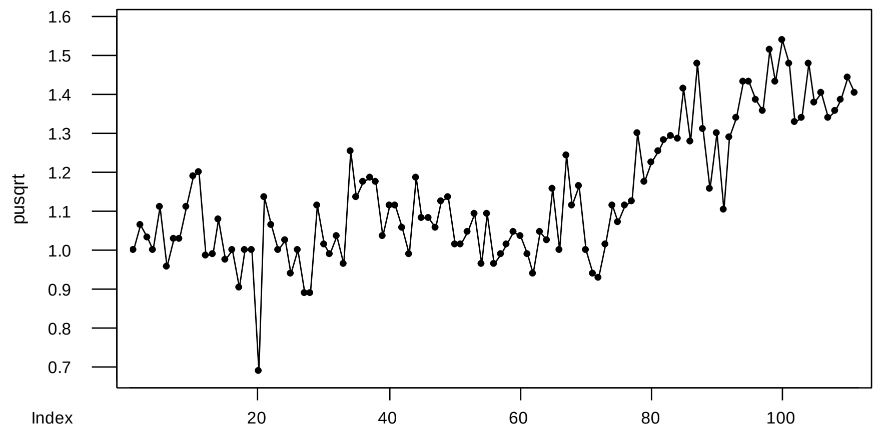


Fig. 4: Time plot of the square root transformed data

3.2. Descriptive and ARIMA models for transformed series

3.2.1Descriptive Model:

Time plot of the original series (Y_t) is shown as Figure 1 above indicates the presence of trend. Two suggested trend lines (linear and quadratic) were fitted and shown in Figures 5 and 6 respectively. Their respective accuracy measures MAD, MSD and MAPE were examined for adequacy of fit.

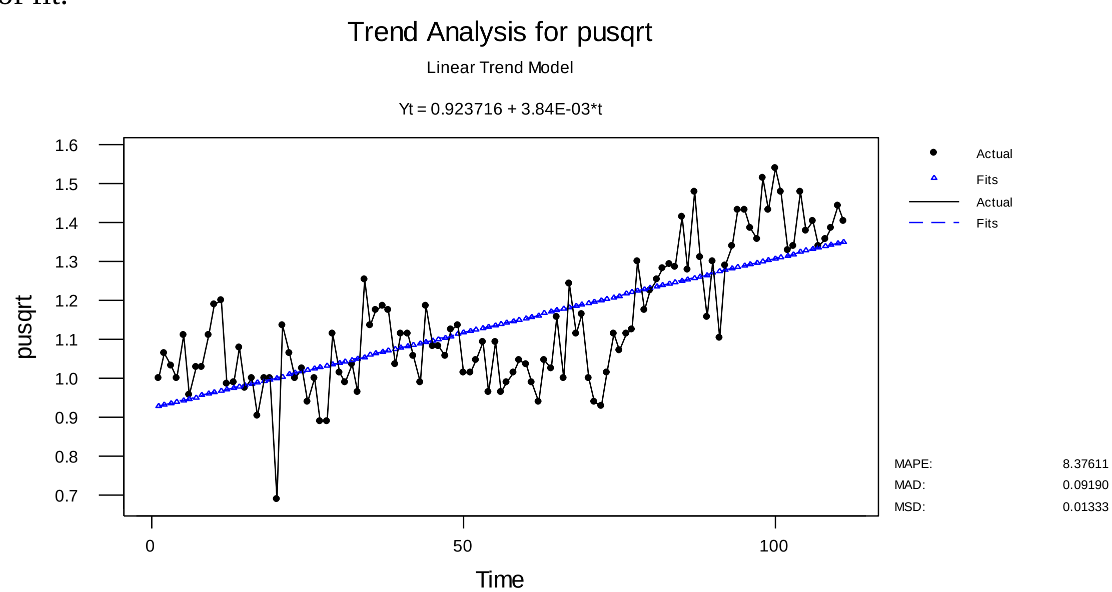


Fig. 5: Linear trend plot of Y_t

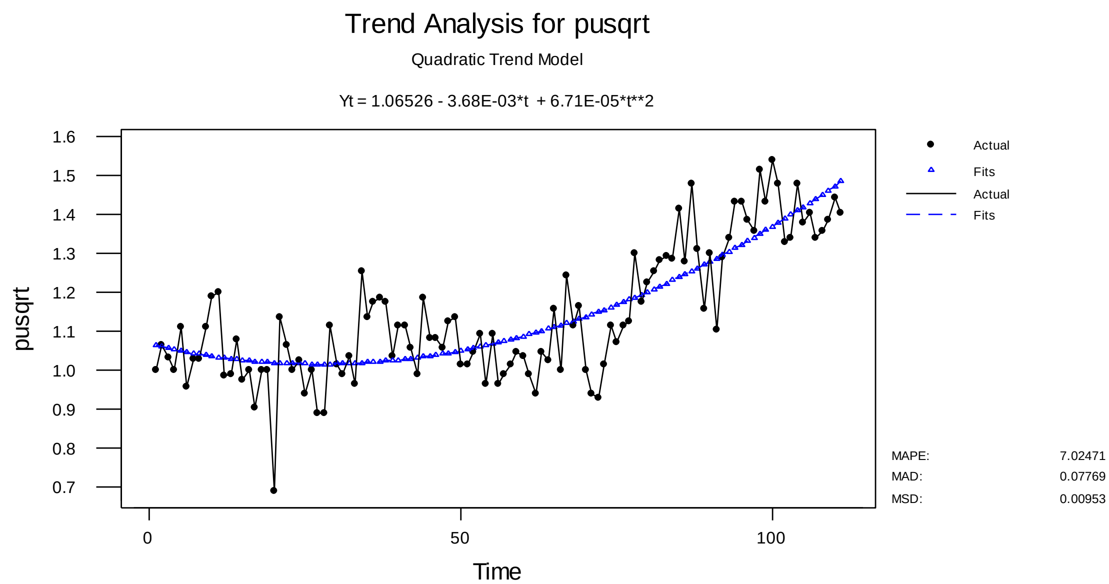


Fig. 6: Quadratic trend plot of Yt

Table 5: Accuracy measures

	Type of trend	
Accuracy measure	Linear	Quadratic
MAPE	8.37611	7.02471
MAD	0.09191	0.07769
MSD	0.01333	0.00953

From Table 5, it is evident that least error is committed while fitting quadratic trend than in fitting linear trend. This points to the fact that quadratic trend is the line of best fit with trend line equation given by

$$Y_t = 1.06526 - 3.68E - 03t + 6.71E - 05t^2 \qquad t = 1, 2, \cdots, 109$$

Successive substitution of values of t in the trend line equation gives the trend values while subtracting the trend values from the original data gives the error component. The plot of the error component from the descriptive model is shown in Figure 7.

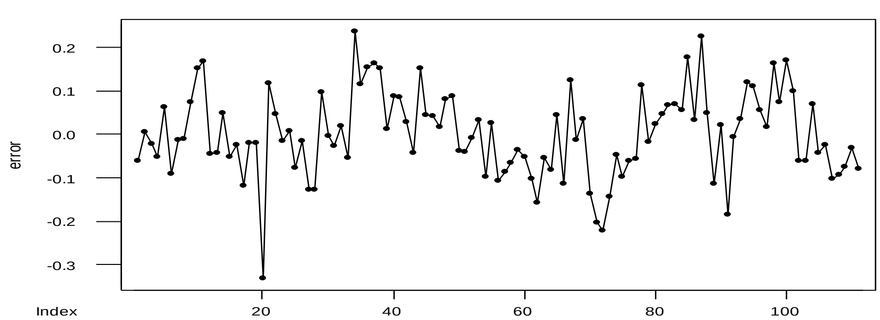


Fig. 7: Plot of the error component.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies

Vol.5, No.1, pp.45-60, February 2017

Published by European Centre for Research Training and Development UK (www.eajournals.org)

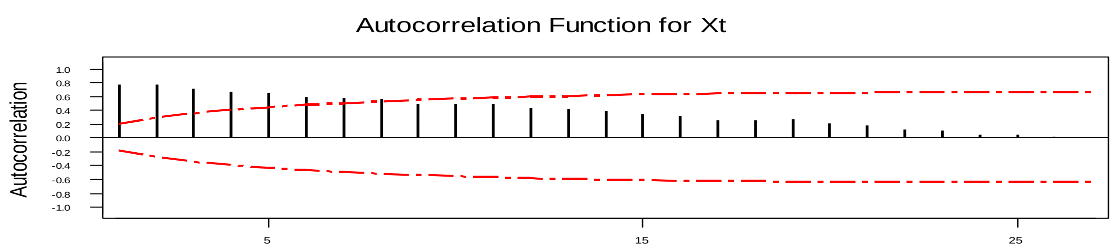
Figure 7 fluctuates along the horizontal line through zero which indicates that the trend has been removed and the series is random.

3.2.2: Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) Model

The time plots of the series is shown in Figure 1, while the corresponding ACFs and PACFs are shown in Table 6 and plots shown in Figures 8 and 9 respectively.

Table 6: ACF and PACF at various lags

Lag (k)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ACF	0.780	0.770	0.720	0.665	0.657	0.592	0.576	0.567	0.497	0.485
PACF	0.780	0.414	0.138	-0.003	0.103	-0.050	0.023	0.091	-0.106	-0.013
Lag (k)	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
ACF	0.489	0,433	0.420	0.382	0.337	0.313	0.258	0.258	0.266	0.212
PACF	0.142	-0,076	-0.047	-0.001	-0.091	-0.052	-0.028	0.044	0.108	-0.054
Lag (k)	**21**	**22**	**23**	**24**	**25**	**26**	**27**			
ACF	0.172	0.121	0.110	0.048	0.040	0.019	0.006			
PACF	- 0.152	-0.123	0.060	-0.126	0.047	0.046	0.006			

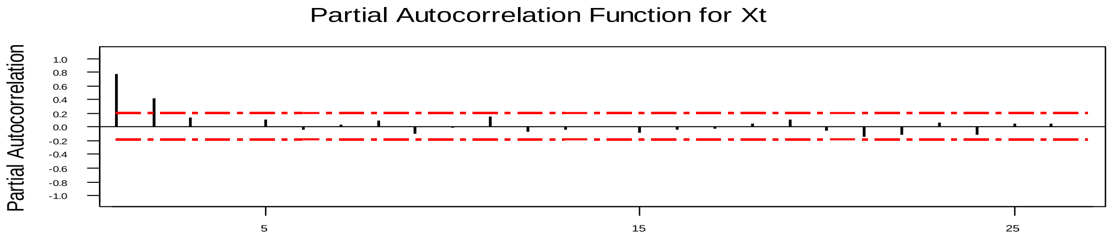


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.78	8.14	68.14	8	0.57	2.15	414.46	15	0.34	1.08	581.47	22	0.12	0.37	633.52
2	0.77	5.40	135.23	9	0.50	1.81	444.35	16	0.31	0.99	594.19	23	0.11	0.34	635.22
3	0.72	4.07	194.35	10	0.48	1.72	473.09	17	0.26	0.81	602.96	24	0.05	0.15	635.55
4	0.66	3.29	245.27	11	0.49	1.69	502.64	18	0.26	0.80	611.79	25	0.04	0.12	635.78
5	0.66	2.97	295.49	12	0.43	1.46	526.06	19	0.27	0.83	621.31	26	0.02	0.06	635.84
6	0.59	2.49	336.69	13	0.42	1.38	548.24	20	0.21	0.65	627.41	27	0.01	0.02	635.84
7	0.58	2.29	376.01	14	0.38	1.24	566.85	21	0.17	0.53	631.48				

Fig. 8: Plot of the ACF of the original series

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Vol.5, No.1, pp.45-60, February 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.78	8.14	8	0.09	0.95	15	-0.09	-0.95	22	-0.12	-1.28
2	0.41	4.32	9	-0.11	-1.11	16	-0.05	-0.54	23	0.06	0.63
3	0.14	1.44	10	-0.01	-0.14	17	-0.03	-0.30	24	-0.13	-1.31
4	-0.00	-0.03	11	0.14	1.48	18	0.04	0.46	25	0.05	0.49
5	0.10	1.08	12	-0.08	-0.79	19	0.11	1.13	26	0.05	0.48
6	-0.05	-0.52	13	-0.05	-0.49	20	-0.05	-0.56	27	0.01	0.07
7	0.02	0.24	14	-0.00	-0.01	21	-0.15	-1.58			

Fig. 9: Plot of the PACF of the original series

The plots of the transformed series show that the series appears to be moving upwards in what appears to be like a curve trend. Table 6 shows that the corresponding autocorrelation function (ACF) of the series decayed slowly from a value of 0.780 at lag 1 to value of 0.006 at lag 27, confirming the presence of trend in the series. This suggests that the series requires differencing to remove the trend. The corresponding partial autocorrelation (PACF)also shown in Table6 has a spike at lag 1and 2 with values 0.780 and 0.412, in other words, it cuts off after lag 2 which shows that second(order)differencing is required to remove the trend. The plot of the 2nd order differenced series is shown in Fig. 10.

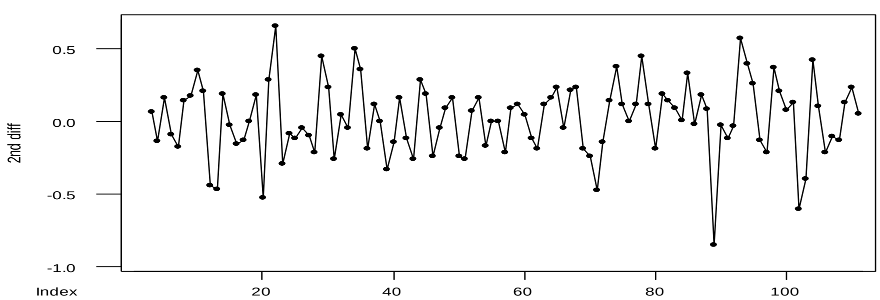


Fig. 10: Time plot of the 2nd order differenced series.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies

Vol.5, No.1, pp.45-60, February 2017

Published by European Centre for Research Training and Development UK (www.eajournals.org)

Table 7: ACF and PACF of the 2nd order differenced series.

Lag (k)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ACF	0.104	-0.307	-0.024	-0.112	0.023	-0.066	-0.003	0.045	-0.197	-0.059
PACF	0.104	-0.321	0.058	-0.244	0.105	-0.249	0.139	-0.174	-0.120	-0.113
Lag (k)	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
ACF	0.142	0.037	0.045	-0.000	-0.010	0.021	-0.108	0.018	0.210	0.121
PACF	0.078	-0.092	0.099	-0.091	0.088	-0.081	-0.024	-0.021	0.221	0.107
Lag (k)	**21**	**22**	**23**	**24**	**25**	**26**				
ACF	-0.154	-0.132	0.033	-0.105	-0.061	0.119				
PACF	-0.098	0.047	-0.026	-0.132	-0.020	0.053				

When the ACFs of the differenced series are compared with the 95% confidence limits; $(\pm 2/\sqrt{n-1} = \pm 0.1907)$, there are spikes at lags 2, 4 and 6 which are sub multiples of seasonal lags. The PACF, on the other hand, appears to have cut-off after lag 2. This confirms that the second order difference was sufficient to achieve stationarity in mean. The plot of the differenced series of Figure 10 shows that the series fluctuates around the horizontal line through zero, which also confirms that the trend has been removed.

The cut-off in PACF after lag 2 tends to suggest that the model to be used is Autoregressive Integrated Moving Average (ARIMA) (p, d, q), with p = 2, d = 2, q = 0. The parameter estimates of the model ARIMA (2, 2, 0) (without constant) given by the MINITAB software is;

Table 8: ARIMA Model estimates

Type	**Coef**	**StDev**	**T_{cal}**
AR 1	-0.7330	0.0768	-9.54
AR 2	-0.6221	0.0767	-8.11

Hence, the fitted model for thesquare root transformed data (Yt) is

$$Yt = -0.7330Y_{(t-1)} - 0.6221\, Y_{(t-2)} + e\,(t) \qquad (4)$$

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Table 9: ACF and PACF of the 2^{nd} order differenced series.

Lag (k)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ACF	- 0.142	- 0.341	-0.141	-0.075	0.289	-0.124	0.072	0.077	-0.177	-0.052
PACF	- 0.142	- 0.369	-0.307	-0.402	-0.049	-0.399	-0.066	-0.070	-0.154	-0.273
Lag (k)	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
ACF	0.122	0.017	0.025	-0.032	0.026	0.007	-0.210	0.060	0.148	0.123
PACF	0.023	- 0.259	-0.082	-0.093	0.085	-0.024	-0.075	-0.147	-0.074	0.065
Lag (k)	**21**	**22**	**23**	**24**	**25**	**26**				
ACF	- 0.123	- 0.120	0.084	-0.092	0.046	0.101				
PACF	- 0.055	0.066	0.153	-0.048	0.005	0.108				

Forecast from the respective models are given in Table 10 below.

Table 10: Points forecast from the models

Lead K	$t_{(t_o+1)}$	Actual	**Descriptive method**		**Probability method**	
			Forecast	Forecast error	Forecast	Forecast error
1	110 (2013/2014	79	70	9	70 points	9
2	111 (2014/2015)	75	74	1	73 points	2

The summary statistics for the purpose of comparison is shown below in Table 11.

Table 11: Summary **statistics** table

Characteristics	Descriptive method	ARIMA model
Mean	1.1387	0.0071757
Median	1.1106	0.18857
Std deviation	0.16944	0.11306
Skewness	0.0023235	0.00017594
Kurtosis	0.0020830	0.00061876
Parameter estimates	A = 1.06526 B = -0.003368 C =0.0000 671	$\hat{\phi}_1$ = -0.7330 $\hat{\phi}_2$ = -0.6221
Error mean	0.00018055	0.000001662
~~MSE~~ (MSD)	0.0095290	0.014667
~~MAE~~(MAD)	0.0077680	0.096183
MAPE	1.87	2.35
Forecast error (average)	5.0	5.5

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

DISCUSSION

It is seen that the periodic variances exhibited instability and according to Bond and Fox [17], a transformation is needed. A summary of Barttlet's transformation as given by Akpanta and Iwueze[12] shows that a square root transformation is most appropriate in the analysis of the data. The plots of the periodic means and periodic standard deviations(against the period) shows that the plots are not moving in the same direction and according to Iwueze and Nwogu[18], the appropriate model to adopt is the additive model. The additive model was fitted to the transformed data for the descriptive part of the work. Evaluation of the study series shows that both descriptive and Box-Jenkins ARIMA models are all effective and efficient methods.

From Table 11 above, accuracy measures (MSE, MAE and MAPE) generated from the residuals of the respective models (methods), show that the estimates from descriptive model is closer to zero than those from the ARIMA model. In view of these, it would seem reasonable to suggest that until other results proves otherwise, that descriptive method is more efficient for the study data than the Box-Jenkins ARIMA model.

The methods adopted in this work is different from multivariate ARIMA method used by Yiannakis et al [10] and Elo-rating used by Haaren Davis [11] and the resultants results vary.

REFERENCES

1. Soar P, Tyler M (2005). The Official illustrated history of Arsenal p. 23.
2. Forbes (2005). Soccer Money league.
3. Soar P, Tyler M (2005). The Official illustrated history of Arsenal p. 32 - 33.
4. Hughes G (2004). Arsenal the Invincibles. BBC Sport 15th May, 2004.
5. Frazer A (2004). Arsenal run ends at 45. BBC sport 25th October, 2004.
6. C Anderson and D Sally (2003). The number game: why everything you know about football is wrong. Penguin, UK.
7. A Joseph, A E Fenton and M Neil (2006): Predicting Football results using Bayesian nets and other machine learning techniques. Knowledge-based systems. Vol. 19 (7) pp. 544 – 553.
8. H Rue and O Salvessen (2000). Prediction and Retrospective analysis of soccer matches in a league. Journal of the Royal statistical society: series D (the Statistician) vol. 49 (3) pp. 399-418.
9. L Addie, M O Amoamah, I W Abdul and O Antwi (2014). Modeling the trend of performance of Manchester United Football club in the 1960-2013 English premiership. Journal of mathematical theory and modelling. Vol 4 (3).
10. A Yiannakis, M J P Selby (2006). Forecasting in sports, the power of social context – A Time Series Analysis with English Premier League soccer. International Review for the sociology of sport: vol. 41 (1) pp. 89-115.
11. J V Haaren, J Davis. Predicting the final league tables of domestic football leagues. 5th Int'l conference on mathematics in sports. pp. 202-207.
12. Akpanta, A. C. and Iwueze, I. S. (2009). *On applying the Bartlett transformation method to time series data*. Journal of Mathematical Sciences, vol. 20 (3), pp. 227-243.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

13. Barttlet, M. S. (1947). The use of Transformations. Biometrica 3: 39-52
14. Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Franscisco: Holden-Day.
15. Wei, W. W. (1989). *Time Series Analysis: Univariate and Multivariate Methods*. Redwood City: Addision-Wesley Publishing Company, Inc.
16. Osborne, J. (2002). *Notes on the use of data transformations*, J. PracticePractical Assessment, Research & Evaluation, 8(6).
17. Bond, T. G. and Fox, C. M. (2001). Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Mahwah, N. J. (Ed), Lawrence Eribaum.
18. Iwueze, I. S. and Nwogu, E. C. (2004). Boys-Ballot estimates for time series decomposition. Global Journal of Mathematical Sciences, Vol. 3, No. 2. pp. 83- 98.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)