St Xavier's College (Autonomous)

Kolkata

Department Of Statistics

A PROJECT WORK ON

TIME SERIES ANALYSIS

NAME: DEBARSHI CHAKRABORTY

ROLL NO: 481

REG NO: A01-1112-0790-17

SUPERVISOR'S NAME: DR.AYAN CHANDRA

I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.

Dchakraborty

Signature of the student

CONTENTS

INTRODUCTION

This project is an application of Time Series Analysis, where we have some secondary data in our hand. We have collected data on the number of goals scored in every season of the UEFA CHAMPIONS LEAGUE over the last 64 years; the data is available for 1955-56 to 2018-19.Clearly,this is a time series data arising from the field of sports. We consider the successive years or rather seasons(like 1955-56,1956-57,...,2018-19) as our time points labelling them as t=1,2,3,....,64 respectively and the total no of goals scored in each season as our variable of interest i.e. $\{X_t\}_{t=1,2,3,...,64}$.The objectives of the analysis and the techniques used, values obtained, etc are explained in detail inside the project.

OBJECTIVE

Now, having this data in hand, our objective is to find out that if the no of goals scored in the successive years over the last 64 years i.e. our variable of interest exhibit any increasing/ decreasing behaviour or follow any particular pattern in terms of trend, seasonality, cyclical fluctuations, etc or not. In brief, we are trying to fit an appropriate model which can explain the nature of our variable of interest over time and verify if the model is good enough or not i.e. here we wish to develop a mathematical model which explains the observed pattern of $X_1,X_2,...,X_t$. This model may depend upon unknown parameters which needs to be estimated.

SOURCE OF DATA

The data used in this project is obtained from the following website-

https://www.worldfootball.net/stats/champions-league/1/

ANALYSIS AND TECHNIQUE USED

At first we note that, one problem to directly work with this raw data is that over the years the format of the tournament and consequently the **number of matches played** have changed. Therefore, it **is not meaningful / worthwhile to work with the number of goals** scored as it is directly proportional to the no of matches played. So, we will take into account the goal ratio in our context which is defined as

G.R =(no of goals scored in a particular season)/(no of matches played in that season)

Now, our new variable of interest i.e the Goal Ratio is a ***discrete time point continuous state space*** time series data and we can carry out a suitable analysis with this. Moreover, an extra advantage in working with our new variable of interest say $\{X_t\}_{t=1,2,3,...,64}$ is that after estimation ,we can easily revert back to our original variable of interest i.e the no of goals just by multiplying the estimates by the no of matches played.

The first thing to do in time series analysis is to plot the observations against time which is popularly known as the TIME SERIES PLOT. From this graph, we try to notice or observe certain important features like trend, seasonality, cyclic patterns, outliers, discontinuities, turning points, etc and try to interpret them subjectively.

Now, the time series data is a combined effect of 4 main factors broadly known as

(i)Trend-T_t

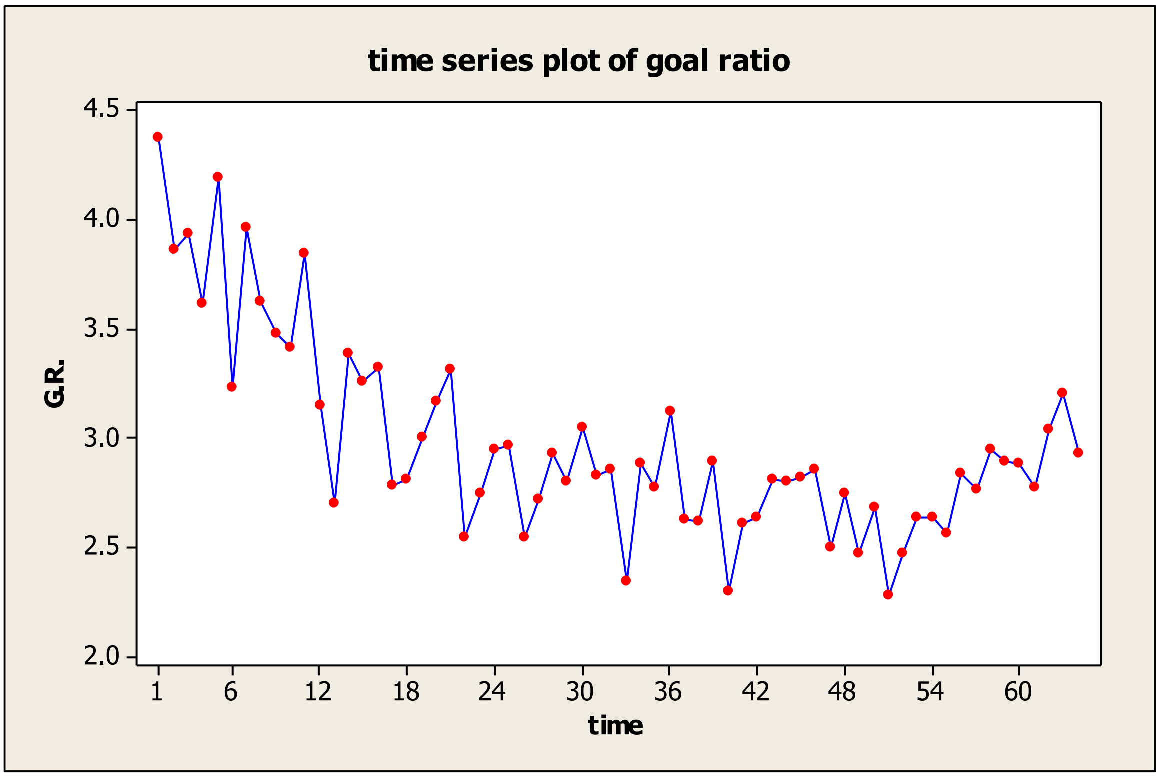
(ii)Seasonality-S_t

(iii)Cyclical-C_t

(iv)Irregular variations-I_t

Attributing a time series due to the four components is called **Classical Decomposition Method.**

Now, at first we plot the graph and try to observe the points mentioned above.



Before observing other features ,we want to state one thing clearly.

→**Justification why seasonal variation is not present:** We note that, the data which is in our hand is an yearly data. Clearly, there is no question of variation arising due to months, weeks, quarters of the year, etc and thus seasonality can be assumed to be absent. Hence, we consider the component S_t to be absent in our model, whatever model we consider shortly.

•**Graphical Method:** From the plot of Goal Ratio vs Year we note or observe the following points-

(i)A trend, most probably quadratic in nature, is present in our data set, we will further analyse it through some objective techniques.

(ii)It seems that cyclical variation may be present, even if in a very small amount-but the periods of oscillations are not uniform.

(iii)No outlier , very distinct discontinuities or sharp turning points can be noticed.

(iv)It seems that the trend and cyclical components do not influence each other i.e. we can examine them independently.

•**Choice of Model:**

From the last point it is quite reasonable to consider a model for explaining $\{X_t\}$ to be-

$$\mathbf{X_t=T_t+C_t+I_t} \qquad (\text{as } S_t=0)$$

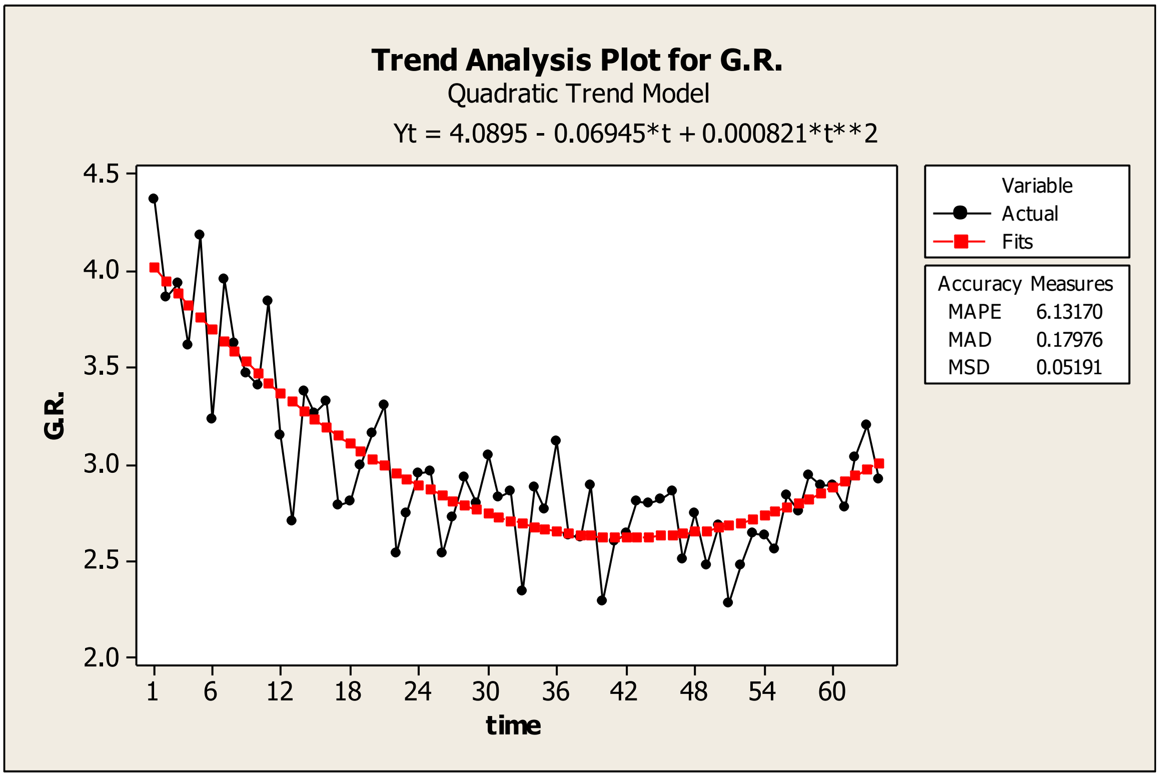
(symbols carrying their usual meanings)

i.e an ADDITIVE MODEL.

•Time series analysis for TREND>>

At first,we will try to estimate the trend present in the data $\{X_t\}_{t=1,2,3,\ldots,64}$.Now,we can either apply the method of moving averages or the method of mathematical curve fitting but as we know that the method of moving averages is a better way usually recommended for removal of seasonality than that of estimating trend values;So,we opt for the method of mathematical curve fitting.

We try to estimate the trend by fitting a quadratic curve.



The trend equation comes out to be-

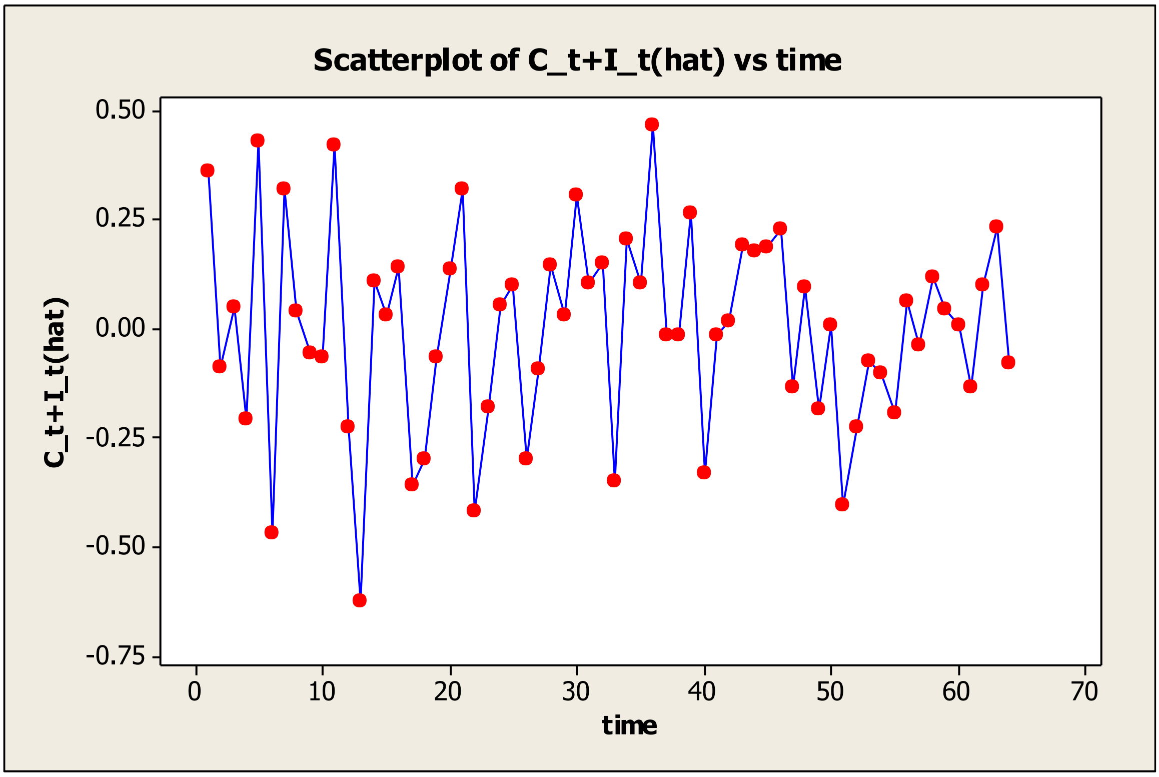
$$\mathbf{T^{\wedge}_t=4.0895-0.06945*t+0.000821*t^2}$$

(where $T^{\wedge}_t$ is the estimated value of trend component at time point t, here $R^2 = 0.75$ i.e. it seems that the fit is moderately good. Moreover, the **MAPE** i.e. Mean absolute percentage error which shows the percentage of absolute deviations of the fitted values from the original time series, **MAD** (Mean absolute error) which reports the absolute magnitude of error regardless of its sign- these measures come out to be quite low in magnitude, therefore our fit seems to be a reasonable one.)

Now, simply subtracting T^$_t$ column from the X$_t$ column we should get the values of **C$_t$ + I$_t$** theoretically.

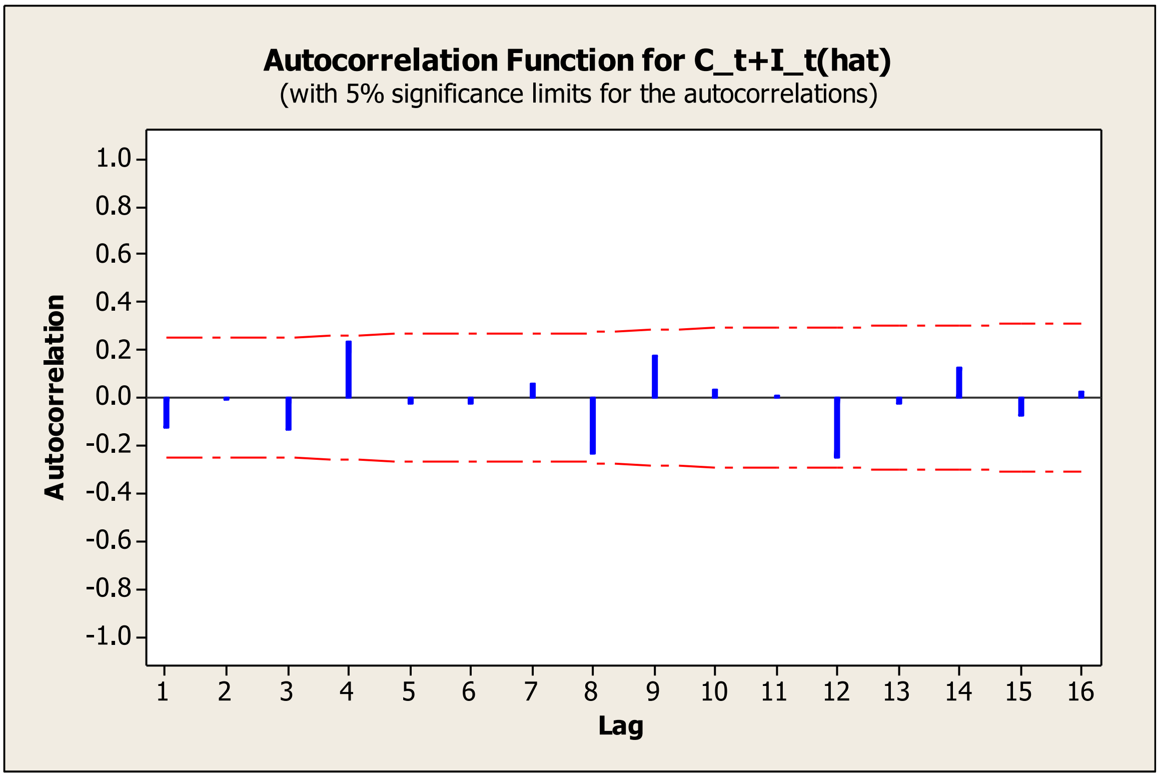
•<u>Time series analysis for CYCLICAL COMPONENT>></u>

The graph of time vs **C$_t$ + I$_t$** (= Y$_t$, say) looks like-



We will now try to estimate the cyclical variation present in our data set and for that we opt to use BOX JENKINS ARIMA METHODOLOGY.

Now, we plot the ACF of Y$_t$ against lag K.



Taking a cue from the Sample Autocorrelation Function, we fit an ARIMA(0,1,1) or IMA(Integrated Moving Average) model to this data-

$$\mathbf{(1\text{-}B)Y_t = (1\text{-}\beta_1 B)e_t}$$

(where B is the backward shift operator and e_t's are white noise)

Let, the fits come out as $Y_1, Y_2, \ldots., Y_{63}$.We lose a point at the beginning while fitting the moving average. The estimated value of the only parameter of the model comes out to be

$\beta^{\wedge}$=0.967305.

Now, the estimated values of these Y_t's is nothing but the estimates of our **cyclical component i.e $C_t^{\wedge}$.**

Now, it is desirable to assess that whether the residuals are actually stationary or not. To investigate this matter in an objective manner, we use the Ljung-Box Test Statistic. The p-values come out to be (corresponding to different choices of lags)-

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi square(observed)	14.6	26.1	39.3	48.9
Chi square(critical)	19.7	35.2	49.8	64
Degrees of freedom	11	23	35	47
p-value	0.201	0.294	0.284	0.396

As all the p-values are greater than our desired level of significance α=0.05 and the observed values of the test statistics is less than the critical value in each case,
Therefore , the null hypothesis H_0:The residuals are stationary(white noise) is accepted.

Moreover, we can check that whether the residuals are random or not. To investigate this matter in an objective manner, we use the large sample approximation of the **Run's Test.** The theory and R-code is discussed later. Here, we briefly note that our null hypothesis *H_0: The data e_t is random* is accepted(as P value=0.858).

•**FINAL ESTIMATE AND CHECK FOR GOODNESS OF FIT>>**

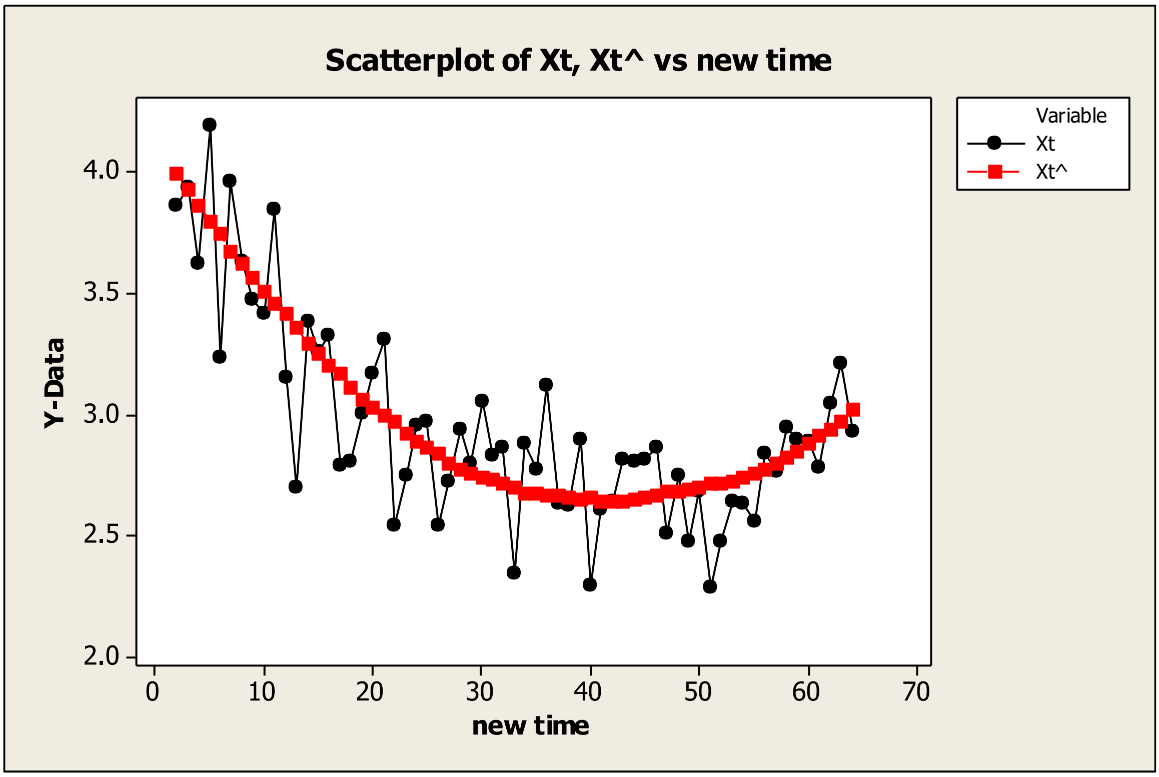
Therefore, we get the values of our fitted model with the help our estimates of deterministic parts just by adding the 2 columns as follows-

$$X_t^{\wedge}=T_t^{\wedge}+C_t^{\wedge}$$

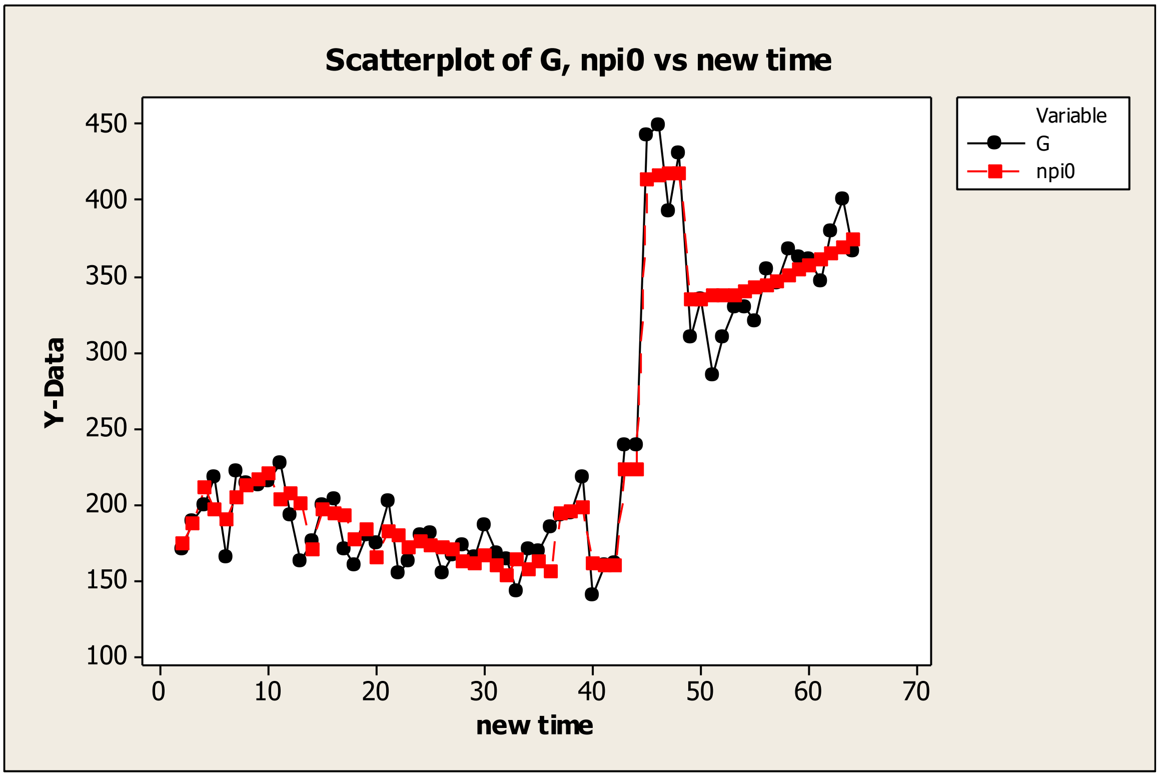
Moreover, we carry out a test to check whether our fit is good enough with the help of a Pearsonian Chisquare statistic which is a nonparametric inferential procedure.

Before that, we provide a graph of the observed goal ratio X_t and estimated goal ratio $X_t^{\wedge}$

To get a visual impression about the fit.



Moreover,we provide a graph of the observed number of goals and estimated number of goals in order to get a visual impression about the nature of the fit.



Here,we amalgamate/combine some classes and then perform our test.

Let T be our test statistic defined as-

$$\mathbf{T=\sum\{(X_t- X_t^{\wedge})^2/ X_t^{\wedge}\}\sim\chi^2_{(33-1-4)}}$$

i.e follows a chisquare distribution with degrees of freedom 28 under the null hypothesis H_0.

(as, 4 parameters are estimated, 1 while estimating cyclical component and 3 while fitting quadratic trend)

Our **T_obs** comes out to be **21.0147** and **T_crit=31.39088(at α=0.3)**,here we take **k=33 and r=4** as while estimating trend we depend on 3 parameters and while estimating the cyclical component we depend on 1 parameter.

Therefore,H_0 in the goodness of fit test is accepted.

[actual frequency=observed no of goals(X_t) and expected frequency=expected no of goals($X_t^{\wedge}$)]

RESULTS

Year	**X_t**	**T_t^**	**C_t^**	**X_t^**
1957	3.86364	3.95389	0.0406139	3.99450
1958	3.93750	3.88855	0.0363353	3.92488
1959	3.61818	3.82485	0.0367479	3.86159
1960	4.19231	3.76279	0.0287896	3.79158
1961	3.23529	3.70238	0.0418913	3.74427
1962	3.96429	3.64360	0.0252505	3.66886
1963	3.62712	3.58648	0.0349096	3.62139
1964	3.47541	3.53099	0.0350970	3.56609
1965	3.41270	3.47715	0.0321324	3.50928
1966	3.84746	3.42494	0.0289747	3.45392
1967	3.14754	3.37439	0.0418414	3.41623
1968	2.70000	3.32547	0.0330567	3.35853
1969	3.38462	3.27820	0.0115263	3.28972
1970	3.26230	3.23257	0.0146288	3.24719
1971	3.32787	3.18858	0.0151225	3.20370
1972	2.78689	3.14623	0.0191821	3.16541
1973	2.80702	3.10553	0.0068062	3.11234
1974	3.00000	3.06647	-0.0031761	3.06329
1975	3.16364	3.02905	-0.0052455	3.02381
1976	3.31148	2.99328	-0.0006737	2.99260
1977	2.54098	2.95914	0.0097518	2.96889
1978	2.74576	2.92665	-0.0042387	2.92241
1979	2.95082	2.89581	-0.0100143	2.88579

1980	2.96721	2.86660	-0.0078882	2.85871
1981	2.54098	2.83904	-0.0043408	2.83470
1982	2.72131	2.81312	-0.0139437	2.79917
1983	2.93220	2.78884	-0.0164894	2.77235
1984	2.79661	2.76621	-0.0112631	2.75494
1985	3.04918	2.74521	-0.0099008	2.73531
1986	2.83051	2.72586	0.0003610	2.72623
1987	2.85965	2.70816	0.0037705	2.71193
1988	2.34426	2.69209	0.0086002	2.70069
1989	2.88136	2.67767	-0.0030532	2.67462
1990	2.77049	2.66489	0.0037060	2.66860
1991	3.11864	2.65376	0.0070374	2.66079
1992	2.63014	2.64426	0.0220067	2.66627
1993	2.62162	2.63641	0.0208254	2.65724
1994	2.89333	2.63020	0.0196609	2.64986
1995	2.29508	2.62564	0.0276212	2.65326
1996	2.60656	2.62271	0.0159107	2.63862
1997	2.63934	2.62143	0.0148623	2.63629
1998	2.81176	2.62179	0.0149620	2.63676
1999	2.80000	2.62380	0.0206839	2.64448
2000	2.81529	2.62744	0.0257686	2.65321
2001	2.85987	2.63273	0.0310676	2.66380
2002	2.50318	2.63967	0.0374781	2.67714
2003	2.74522	2.64824	0.0317906	2.68003
2004	2.47200	2.65846	0.0339221	2.69238
2005	2.68000	2.67032	0.0267168	2.69703
2006	2.28000	2.68382	0.0261599	2.70998

2007	2.47200	2.69896	0.0121019	2.71106
2008	2.64000	2.71575	0.0042857	2.72004
2009	2.63200	2.73418	0.0016689	2.73585
2010	2.56000	2.75425	-0.0017264	2.75253
2011	2.84000	2.77597	-0.0080211	2.76795
2012	2.76000	2.79933	-0.0056653	2.79366
2013	2.94400	2.82433	-0.0067658	2.81756
2014	2.89600	2.85097	-0.0026319	2.84834
2015	2.88800	2.87925	-0.0010736	2.87818
2016	2.77600	2.90918	-0.0007525	2.90843
2017	3.04000	2.94075	-0.0050823	2.93567
2018	3.20800	2.97397	-0.0016713	2.97229
2019	2.92800	3.00882	0.0060351	3.01486

CONCLUSION

Our main objective was to develop a simple mathematical model to explain the observed patterns of $X_1,X_2,.....,X_T$. Now, at the end, the parameters of our model are estimated and the predicted values are obtained with the help of that model. It is given by-

$$\mathbf{X^{\wedge}{}_t=T_t{}^{\wedge}+C_t{}^{\wedge}} \quad \text{OR} \quad \mathbf{X^{\wedge}{}_t = 4.0895-0.06945*t+0.000821*t^2+C_t{}^{\wedge}}$$

After that, we see that the fit is good enough, therefore we can conclude that the model we developed explains the patterns of the time series very well and it can be further used for forecasting purpose.

APPENDIX

(a)Descriptions of the techniques used:-

(i)Method of Mathematical Curve Fitting:-

Let, x_i (i=1,2,...,n) denote the given time series data. We assume that it is dominated by trend only and consider an additive model given by-

$X_t = T_t + I_t$

Where symbols have their usual meanings. Taking a cue from the scatterplot , we fit a suitable polynomial trend to the given time series data. In our case, we try to fit a quadratic curve i.e. we assume

$\mathbf{x_t = T_t + I_t = a_0 + a_1 t + a_2 t^2 + \varepsilon_t}$

$\mathbf{\sum \varepsilon_t^2 = \sum(x_{t-} a_0 - a_1 t - a_2 t^2)^2}$

We thus get the Normal Equations as-

$\mathbf{\sum x_t = na_0 + a_1 \sum t + a_2 \sum t^2}$

$\mathbf{\sum tx_t = a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3}$

$\mathbf{\sum t^2 x_t = a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4}$

By relabeling t as $t = t - (\sum t_i)/n$ i.e. suitable change of origin we have $\sum t = \sum t^3 = 0$

Hence, our normal equations get reduced to –

$\mathbf{\sum x_t = na_0 + a_2 \sum t^2}$

$\mathbf{\sum tx_t = a_1 \sum t^2}$

$\mathbf{\sum t^2 x_t = a_0 \sum t^2 + a_2 \sum t^4}$

By solving these 3 equations simultaneously we get the estimates of the model parameters as-

$\mathbf{a^\wedge_1 =(\sum t x_t)/(\sum t^2)}$

$\mathbf{a^\wedge_2 =\{(\sum t^2 \sum x_t)-n(\sum t^2 x_t)\}/\{(\sum t^2)^2-n\sum t^4\}}$

$\mathbf{a^\wedge_0 =(\sum x_t- a^\wedge_2 \sum t^2)/n}$

Thus, $\mathbf{T_t^\wedge= x_t- a^\wedge_0- a^\wedge_1 t – a^\wedge_2 t^2}$ and $\mathbf{I_t^\wedge=x_t - T_t^\wedge}$

(ii)Box Jenkins ARIMA methodology for estimation of cyclical component:-

Suppose we have a de-trended and de-seasonalised time series data in our hand which still exhibits some non-stationarity . Then, we follow the steps given below to estimate the cyclical component-

→Plot the series, obtain ACF(Autocorrelation Function) and PACF(Partial Autocorrelation Function).
→If the process is stationary, go for model building process, otherwise use prior transformations and differencing to obtain stationarity.
→Examine the ACF and PACF to identify potential models.
→Estimate the parameters of the model and obtain the fitted values.
→Check the ACF/PACF of residuals, perform Portmanteau test of residuals and see whether they are white noise.

Obviously, the stationary process resulting from a properly differenced homogeneous nonstationary series is not necessarily white noise. More generally, the differenced series $(1-B)^d x_t$ (where B is the backward shift operator and x_t is our nonstationary time series data),follows a general stationary Autoregressive Moving Average[ARMA(p,q)] process. Thus, we have our general Autoregressive Integrated Moving Average Model (ARIMA(p,d,q)) as-

$$\boldsymbol{\varphi_p(B)(1-B)^d x_t = \theta_0 + \theta_q(B)e_t} \text{ -----------------------(i)}$$

[where stationary AR operator $\varphi_p(B)=(1-\alpha_1B-...-\alpha_pB^p)$ and the invertible MA operator $\theta_q(B)=(1-\beta_1B-...-\beta_qB^q)$ share no common factors.]

Now, in our problem ,we will discuss only about the ARIMA(0,1,1) or IMA(1,1) model-

When p=0,d=1 and q=1, the model (i) becomes-

$$\boldsymbol{(1-B)x_t = (1-\beta_1B)e_t} \text{ ---------------------------------(ii)}$$

(where $-1<\beta_1<1$)

This IMA model is reduced to a stationary MA(1) model after taking the first difference i.e if y_t denotes the diferrenced time series then

$y_t= e_t+\beta_1e_{t-1}$ (where e_t is a purely random process)

Now, for MA(1) process we have $\rho(1)=\beta_1/(1+\beta_1^2)$; $|\beta_1|<1$

Let, $y_1,y_2,....,y_n$ denote the stationary data then the estimate of β_1 can be obtained by solving the quadratic equation-

$\mathbf{r_1= \beta_1^{\wedge}/(1+(\beta_1^{\wedge})^2)}$

where $r_1=\{\sum_{t=1}^{n-1} (y_t-y)(y_{t+1}-y)\}/\{\sum_{t=1}^{n} (y_t-y)^2\}$ (where $y=(y_1+y_2+..+y_n)/n$)

To ensure invertibility , we choose that value of $\beta_1^{\wedge}$ for which $|\beta_1^{\wedge}|<1$.

Thus, we get the estimate of our model parameter of the model (ii) i.e our IMA(1,1) or ARIMA(0,1,1) model.

After fitting the model, we are to check if the residuals are purely random or not, although we can get a rough idea of that just by looking at the correlogram,we can perform some test to examine the matter objectively.

Ljung Box Test:-

This test may be defined as-

H_0: The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

H_1: The data are not independently distributed; they exhibit serial correlation.

The test statistic is given by-

$Q^*=n(n+2)\sum_{i=1}^{k} \{\rho_i^2 *(n-i)^{-1}\}$

(where n is the no of sample observations and ρ_i is the autocorrelation coefficient at lag i(i=1(|)k))

This statistic has a distribution very close to a χ^2 distribution with degrees of freedom (k-m) where m is the number of parameters in the model which has been fitted to the data.

It is normal to conclude that the data are not white noise if the value of Q lies in the extreme 5% of the right hand tail of the $\boldsymbol{\chi^2_{(k-m)}}$ distribution i.e

$Q_{obs} > \chi^2_{1-\alpha;(k-m)}$,α being the level of significance.

(iii)One Sample Runs Test(Test for randomness):-

Let $X_1,X_2,...,X_n$ be a set of observations drawn from a continuous distribution. Here, we are to test-

H_0: observations arise from a random process against H_1:observations arise from a non-random process.

TEST PROCEDURE:

Let, the given observations are $x_1,x_2,...,x_n$ .We arrange them in ascending order of magnitude. Let $\mathbf{x_{me}}$ denote the median of the sample observations. Now, to the original set of observations, we assign the value 0 if it is less than $\mathbf{x_{me}}$ and the value 1 if it is greater than $\mathbf{x_{me}}$. Thus, the original set of observations is now transformed into a sequence of 0's and 1's.We define our test statistic as below-

W=total number of runs=number of 0 runs + number of 1 runs

Under H_0,an observation from the given set is as likely to be replaced by 0 as it is to be replaced by 1.Hence,as our alternative is simply non-randomness, too few or too large number of runs will indicate departure from the null hypothesis H_0.As our sample size is large enough, we can use large sample approximation for our case particularly.

It can be shown that , under H_0 ,

E(W)=1+(n/2) ; n is even **V(W)=n(n-2)/4(n-1) ; n is even**

=1+(n-1)/2;n is odd **=(n-1)(n-3)/4(n-2);n is odd**

Then, we have our test statistic as- **$Z= \{W-E_{H0}(W)\}/\sqrt{V_{H0}(W)}$**

Therefore , clearly **$Z \sim N(0,1)$ (asymptotically under H_0)**

Here, we reject H_0 against H_1 at level of significance α iff **$|Z_{obs}| > \tau_{(\alpha/2)}$**

(where $\tau_{(\alpha/2)}$ is the 100(1-α)% point of a N(0,1) distribution)

(iv)Test for Goodness of Fit (using Pearsonian chisquare statistic):-

Here, our problem is to test whether a population distribution is of a specified kind.

Suppose, a population is divided into "k" mutually exclusive and exhaustive classes, p_i being the population proportion of the i^{th} class (i=1,2,..,K).here, the problem is to test-

$H_0:p_i=p_i^0 \ \forall \ i=1(i)K$ against H_1:Not H_0.

(where p_i^0s are specified values of p_i s)

Suppose, a random sample f size n is drawn from the given population and let f_i be the number of sample members belonging to the i^{th} class(i=1(1)K).The appropriate test statistic is then given by-

$$\mathbf{T=\sum\{(f_i-np_i^0)^2/np_i^0\}} \sim \chi^2_{(K-1)} \text{ (asymptotically under } H_0)$$

Now, greater the difference between the observed frequency f_i and the expected frequency np_i^0 (under H_0), greater will be the value of T. Hence, a very high value of T will lead to the rejection of H_0.Thus we will be rejecting H_0 against H_1 at level of significance α iff

$$T_{obs} > \chi^2_{\alpha;(K-1)}$$

(where $\chi^2_{\alpha;(K-1)}$ denotes the 100(1-α)% point of the $\chi^2_{(K-1)}$ distribution)

In order that the fit maybe good, it is necessary that H_0 be accepted at a high level of significance ,say $\alpha=0.3$,rather than $\alpha=0.05$ or 0.01.

NOTE:

It may happen that the population distribution depends on some unknown parameters. Hence, in this case,p_i^0s will not be completely specified by H_0 but will depend on unknown parameters. Let, there be “r” such parameters where r<K-1.Also suppose that, these parameters are estimated on the basis of sample values and let the corresponding estimate of p_i^0 be $p{\wedge}_i^0$(i=1(1)K).The modified test statistic will then be given by-

$$\mathbf{T=\sum\{(f_i-np{\wedge}_i^0)^2/np{\wedge}_i^0\}} \sim \chi^2_{\mathbf{(K-1-r)}} \text{ (asymptotically under } H_0)$$

(provided that the estimated expected frequencies are large enough so that chi square approximation remains valid, otherwise we have to amalgamate more than one classes to make each of the expected frequencies atleast 5)

(b) R code (for one sample Runs Test):-

```
rm(list=ls())

x=c(0.358438,-0.448690,..........................,-0.314856)

z=summary(x)

median(x)=z[3]

n=length(x)

p=array(0)

for(i in 1:n)

{

if(x[i]>median(x))

p[i]=0
else

p[i]=1
}
p
w=0
for(i in 1:(n-1))
{
if(p[i+1]>p[i] | p[i+1]<p[i])
w=w+1
}
1+w
Ew=1+(n/2)
Vw=(n*(n-2))/((n-1)*4)
t=(w+1-Ew)/sqrt(Vw)
t1=abs(t)
t1
qnorm (0.975)
```

BIBLIOGRAPHY

1) Applied Time Series Analysis and Forecasting

-T.M.J.A Cooray

2) The Analysis Of Time Series-An Introduction

-Chris Chatfield

3) Fundamentals Of Statistics(Volume 1)

-A.M. Gun

M.K. Gupta

B. Dasgupta

SOFTWARES USED

1) Minitab 16

2) R (i386 3.2.4/x64 3.2.4) Revised

OTHER REFERENCES

Understanding Ljung Box test procedure:
https://en.wikipedia.org/wiki/Ljung%E2%80%93Box_test

ACKNOWLEDGEMENTS

At last, I want to say that whatever content, creativity I have tried to put in this project work is not just a result of my efforts alone; I couldn't have completed it in such a manner without the help of the professors of our department who played the most important role to enrich my knowledge in the subject throughout the 3 years. I specially want to thank my project guide and supervisor Dr. Ayan Chandra Sir who always encouraged me from the day I started this work, cleared my each and every doubt and provided me with all sorts of references and support I needed in a very friendly manner. I am really grateful to him and consider myself fortunate to work under a guide and such a nice person like him.