**Project Report**

**Time Series Analysis and Forecasting Software**

**Group 20**

**Group Members (Name and Roll number):**

**Meenakshi Rawat – 145280015**

**Shreya Singh – 145280020**

**Jaydeep Sengupta – 145280025**

**Soumyajeet Nag - 145280030**

**Guided by**

**Chandra Maloo**

**Binoy Gupta**

**Introduction:**

The aim of the project is to forecast trend of the data given by the user using Time Series Analysis and forecasting of GDP in C++.The reason of picking up this project is to integrate our knowledge of Statistics along with coding in C++ that can be used to predict future behavior of the data that the user inputs by fitting curve of the data under consideration making use of graphs. In Time Series Analysis by studying the type and nature of variations in the data relating to prices of commodities, sales and profits of stocks, agricultural and industrial production etc. with time, prediction is made about nature of variation of the data in future. Since the variations may occur either during a long of period of time or during a small period of time, we are using different methods like Moving Average Method and Ratio to Moving Average Method to study these variations**.**

**Purpose:**

The project is aimed at designing a Program using C++ which is used to predict the variation in the data given by the user in future by analyzing the data using Statistical Methods. The user is required to give the data which will contain two variables-one of them being time and the other being the variable whose forecasting is to be done.

**Division of Work:**

1. SRS and Group Diary- Shreya
2. Code for Moving Average and seasonal components – Meenakshi
3. Code for Linear, Exponential and Quadratic Curve

fitting –Soumyajeet and Jaydeep

1. Code for Error Computation-Shreya and Jaydeep
2. Code for plotting graph- Meenakshi and Soumyajeet
3. Main function- Shreya
4. User Manual-Soumyajeet and Meenakshi
5. Project Report- Soumyajeet ,Jaydeep and Shreya

**Description:**

**Data:**

Here we shall deal with the statistical data which relate to successive intervals or points of time. User is required to input the data that will contain two variables one will time and the other being the variable whose future value we have to forecast which could be -

• Economic - Price of a commodity, monthly data for unemployment, hospital admissions, population of a country etc.

• Financial - Daily exchange rate, a share price, etc.

• Environmental - Daily rainfall, air quality readings.

**Components of time series:**

Time series data contains –

* **Trend Component**:

Trend is a long term movement in a time series. It is the underlying direction (an upward or downward tendency) and rate of change in a time series, when allowance has been made for the other components.

A simple way of detecting trend in seasonal data is to take averages over a certain period. If these averages change with time we can say that there is evidence of a trend in the series. There are also more formal tests to enable detection of trend in time series. It can be helpful to model trend using straight lines, polynomials, Exponential etc. In the project we are to going to use following methods to measure trend components –

(i**) Method of Moving average.**

(ii) **Method of Mathematical curves (polynomial and exponential).**

* **Cyclical Component**:

In weekly or monthly data, the cyclical component describes any regular fluctuations. It is a non-seasonal component which varies in a recognizable cycle.

* **Seasonal Component**:

In weekly or monthly data, the seasonal component, often referred to as seasonality describes any regular fluctuations with a period of less than one year. For example, the costs of various types of fruits and vegetables, unemployment figures and average daily rainfall, all show marked seasonal variation. In the project we will try to compare the seasonal effects within the years, from year to year; removing seasonal effects so that the time series is easier to cope with; and, will also try to adjust a series for seasonal effects using Ratio to Moving average method to measure Seasonal component.

* **Irregular/ Error Component**:

The irregular component is that left over when the other components of the series (trend, seasonal and cyclical) have been accounted for.

* **Forecasting**:

This is the process of making statements about events whose actual outcomes (typically) have not yet been observed. A commonplace example might be [estimation](http://en.wikipedia.org/wiki/Approximation) of some variable of interest at some specified future date. [Prediction](http://en.wikipedia.org/wiki/Prediction) is a similar, but more general term. Both might refer to formal statistical methods employing [time series](http://en.wikipedia.org/wiki/Time_series), [cross-sectional](http://en.wikipedia.org/wiki/Cross-sectional_data) or [longitudinal](http://en.wikipedia.org/wiki/Longitudinal_study) data, or alternatively to less formal judgmental methods. Usage can differ between areas of application: for example, in [hydrology](http://en.wikipedia.org/wiki/Hydrology), the terms "forecast" and "forecasting" are sometimes reserved for estimates of values at certain specific [future](http://en.wikipedia.org/wiki/Future) times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period.

[Risk](http://en.wikipedia.org/wiki/Risk) and [uncertainty](http://en.wikipedia.org/wiki/Uncertainty) are central to forecasting and prediction; it is generally considered good practice to indicate the degree of uncertainty attaching to forecasts. In any case, the data must be up to date in order for the forecast to be as accurate as possible.

**Outline:**

Themain objective of time series analysis is to forecast for future purpose. To do so we need to select the appropriate model to make the analysis most efficient. In this situation we need

* Components of a time series like trend, error etc.
* Estimating and removing the trend and seasonal components.

In our analysis of time series we assume the multiplicative model i.e. the various components of time series operate proportionately to the general level of the series. According to the multiplicative model,a time series can be expressed as :

Yt=Tt \* St \*Ct \*Et

The error component is assumed to act independently to follow log normal distribution. In the study of trend component we are considering three types of trend : linear, quadratic and exponential.

**Implementation and Challenges:**

The main use of time series analysis is in business and finance sector, weather forecasting, economy and planning structure of a state, also in budget allocation for any real life problem. Realization of the fact that "Time is Money” in business activities, the dynamic decision technologies presented here, have been a necessary tool for applying to a wide range of managerial decisions successfully where time and money are directly related.

In real life situation there are lots of outliers and it is unlikely that one can totally explain a given time series. In forecasting process we try to predict the data within a given range of error. Especially in micro-economic and macro-economic type data we often have to predict for the immediate future. However it may often happen that due to some extraneous effect there is a sudden change of data. The irregularity components often overpower the trend and seasonal components. For example for analyzing data of a stock exchange of a particular day trend has no use. We have to consider other simultaneous changes say situation of international trading market, crude oil price, political stability etc.

**In the project we have used functions:**

Given a bivariate data set of size n (time (t), time series data(Y))

**Trend value:**

These functions will calculate the trend from the given data set. Then we will subtract the calculated trend from Y, and get the seasonal and irregular component.

**void trend( )**

the function will be called to calculate the trend value and will contain three functions-

1. **void linear ( )**

The function estimates the constants and fits a degree one polynomial using formula

**Y= a + b\*t**

wherea and b are constants

1. **void quadratic ( )**

The function estimates the constants and fits a degree two polynomial using formula

**Y = a+ b\*t +c\* (t^2)**

where a, b and c are constants

1. **void exponential ( )**

The function estimates the constants and fits a degree one polynomial using formula

**Y = a + b^ t**

Solved as

log Y = log a + t\* log b

Constants a and b are calculated and then put in the formula Y = a+ b^ t

From the three functions whichever will give least error that value will be considered as trend value

**Seasonal Components:**

In this function we are calculating the seasonal components. We have used ratio to moving average method for this purpose. Then subtracting this from the remaining part we can obtain the error part.

**void seasonal (float [ ],float [ ],int ,int , int )**

the function contains array of observations, total number of observations,total number of years and type of data (monthly/ quarterly/ half yearly/ yearly) it calls the movingaverage function which gives the moving average values and then prints the appropriate calculated seasonal indices

**void movingaverage (float [ ],float [ ], int , int )**

the function contains an array of observations, an empty array which will get values of calculated moving averages, type of data and total number of observations. It calculates moving average depending on type of data user inputs which is basically averaging data depending on type of data using appropriate methods.

**Error Component:**

The remaining part will be the error component which is calculated using function

**void error( [ ] )**

**Graph Plotting:**

We are using files in C++ for plotting graph which will plot a line graph for values of Y corresponding to time t

## References

\*Fundamentals of Applied Statistics by S.C.Gupta and V.K.Kapoor.

\*Articles on time series in Wikipedia

\* [www.stat.duke.edu/~mw/data-sets/ts\_data/gdp](http://www.stat.duke.edu/~mw/data-sets/ts_data/gdp)

\* www. cplusplus.com