Chapter 4: Introduction to Time Series

4.0 Cross-Section Data versus Time-Series Data

Based on the time over which the data are collected, data can be classified as either cross-section data or time-series data.

Cross-section data
Data collected on different elements at the same point in time or for the same period of time.
Example: Total population of each state of Malaysia in year 2000.

Time-series data
Data collected on the same element for the same variable at different points in time or for different periods of time.
Example: New life insurance policies purchased between 1995 and 2000.

Time series are used in various fields such as:
(a) **Economics** time series. Examples include share price on successive days, weekly bank interest rates, export total in successive months, average income in successive months, company profit in successive years, and so on.
(b) **Physical** time series, e.g. meteorology, marine science and geophysics. Examples are wind speed in successive hours, rainfall on successive days, and air temperature measured in successive hours, days or months.
(c) **Marketing** time series. The analysis of sales figures in successive weeks or months is an important problem in commerce.
(d) **Demographic** time series. An example is the population of a country in successive years.

4.1 Stochastic process. Discrete and continuous time series

A **stochastic process** is a collection of random variables that are ordered in time and defined at a set of time points, which may be **continuous** or **discrete**. Examples include the length of a queue, the number of accidents in a particular town in successive months and the air temperature at a particular site on successive days.

A ***discrete stochastic*** *process* is a family of random variables structured as a *sequence* (finite or infinite) and has a discrete time index, denoted Y_t. Discrete stochastic processes may model, for instance, the recorded daily high temperatures in Melbourne, Australia.

A ***continuous stochastic*** process is also a family of random variables but is indexed by a continuous variable and denoted as $Y(t)$. A commonly encountered continuous process is the *Wiener Process*, which describing a particle's position as a function of time as it floats on the surface of a liquid (Brownian Motion). Another commonly encountered continuous process is the *Poisson Process*.

Example:

Year	2000	2001	2002	2003	2004
Sales	75.3	74.2	78.5	79.7	80.2

A time series is said to be **deterministic** if it can be forecast exactly. For **stochastic time series**, the exact prediction is impossible and must be replaced by the idea that future values have a probability distribution which is conditioned by knowledge of the past values.

4.2 Objective of time-series analysis

The purposes of time series analysis are

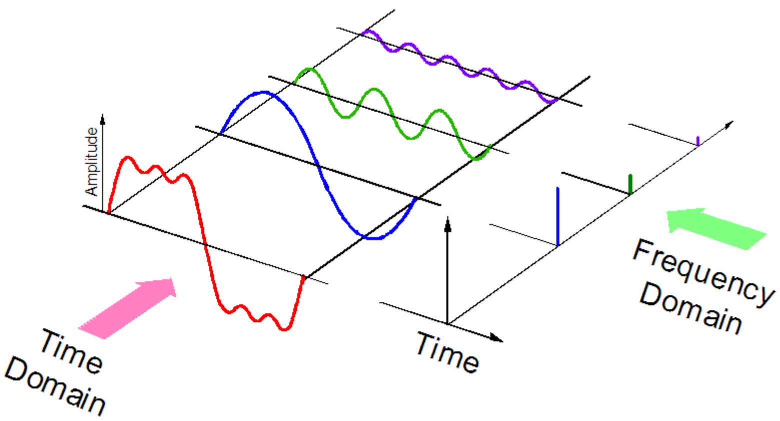
i. to understand or model the behaviour of the observed series, and
ii. to predict or forecast future of a series based on the past values of the series, or
iii. to control the 'quality' of a process in the statistical quality control.

4.3 Time domain and frequency domain

The variation in time of environmental quantities can be studied using the rich branch of statistics known as time series analysis. The main aims of time series analysis are to explore and extract signals (patterns) contained in time series, to make forecasts (i.e. future predictions in time), and to use this knowledge to optimally control processes. The two main approaches used in time series analysis are:

1. **Time domain**
   The inference is made based on autocorrelation function. Time series directly as functions of time. Every point on the time domain plot represents the amplitude at a particular time. There are two basic types of "time domain" models:
   a. **Auto correlation**. Ordinary least square regression models that use time indices as x-variables. These can be helpful for an initial description of the data and form the basis of several simple forecasting methods.
   b. **Cross correlation**. Models that relate the present value of a series to past values and past prediction errors - these are called ARIMA models (for Autoregressive Integrated Moving Average).

2. **Frequency domain**
   The inference is made based on the spectral density function. Every point on a frequency spectrum represents the power or amount of energy at that frequency over a finite time window. There are two basic types of "frequency domain" models:
   a. **Spectral analysis**. Fourier transform decomposes a time series into its frequency components. Fourier transform will only give information on which frequencies are present but will give no information on when they occur. Different time series would have different coefficients of sine and cosine terms and thus different time series can be compared by comparing coefficients.
   b. **Wavelet analysis**. Wavelet transformation analyse the signal in a time scale plane. It contains information on both the time location and frequency of a signal.

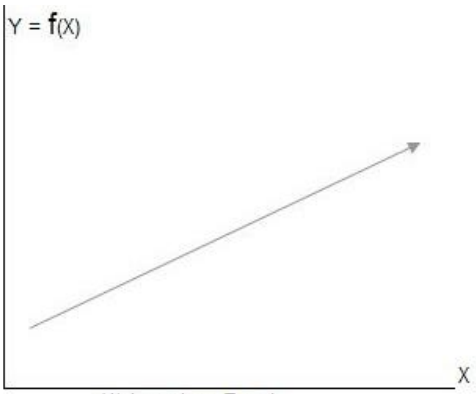


The time domain approach considers regression on past values of the time series; while the frequency domain approach considers regression on sinusoids.

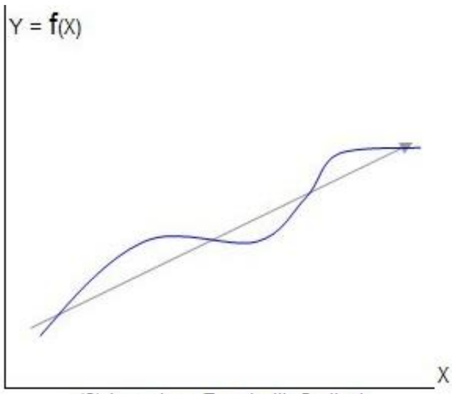
4.4 Types of variation

Traditional methods of time-series analysis are mainly concerned with decomposing the variation in a series into components representing ***trend***, ***seasonal variation*** and other ***cyclic*** changes. Any remaining variation is attributed to '***irregular***' fluctuations.

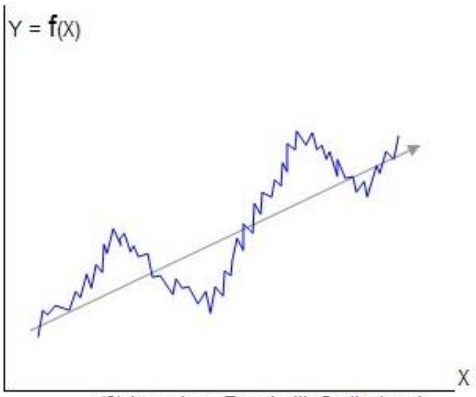
1. **Trend.** The trend of a time series is the underlying long-term movement or tendency of the data. The duration of trend is more than one year, and its fluctuation is due to factors which change slowly over a long stretch of time. The trend does not always show a linear pattern. However, the trend in time series is, in general, represented by a smooth graph. For example, increase in prices, increase in pollution, increase in the need of wheat, increase in literacy rate, decrease in deaths due to advances in science.
2. **Cyclic Variation.** The cyclical variations are long-term cyclic movement of the data, which has four phases (i) prosperity/peak (ii) recession (iii) trough/depression, and (iv) expansion. The patterns of change occur repetitively over duration of more than one year. The long-term cyclic movement is due to the effect or influence of business or economic conditions which are irregular in length and amplitude. For example, economic crisis.
3. **Seasonal variation.** Seasonal variation is the term used to describe patterns of change that recur over short period of time. It is a short-term cyclic movement of the data. The duration is usually less than one year. 'Season' in this case may mean a period of quarter of a month, or even a day. For example, cost of variation for fruits / vegetables, unemployment figures, average daily rainfall, increase in sale of ice cream in summer, traffic on roads in morning and evening hours, foreign exchange rate etc.
4. **Irregular Fluctuations.** Irregular variations are random variations other than those that can be accounted for by the trend, seasonal, or cyclic variations. The changes occur in an unpredictable manner. Bad weather, illness, earth quick, and war are examples of random factors that may occur at any time of the day.



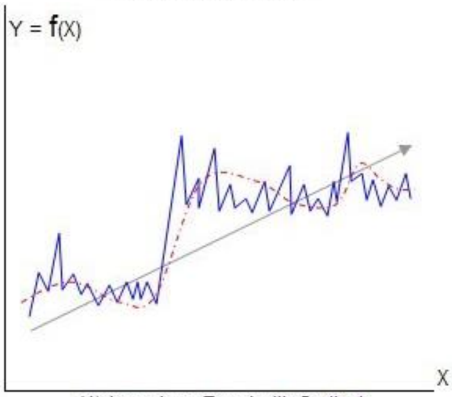
(1) Long-term Trend



(2) Long-term Trend with Cyclical Variations/Movements



(3) Long-term Trend with Cyclical and Seasonal Variations/Movements



(4) Long-term Trend with Cyclical, Seasonal, and Random Variations/Movements

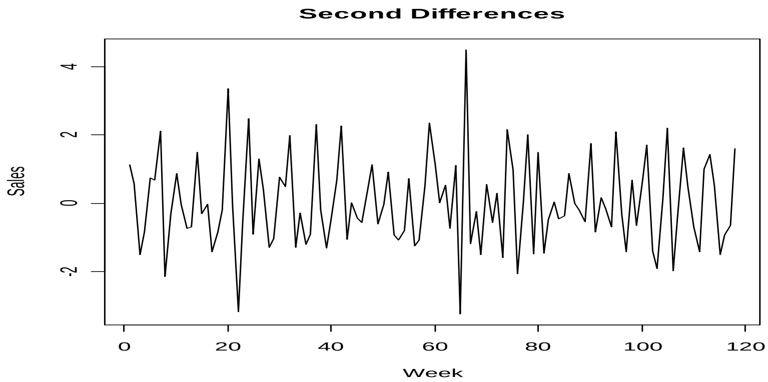
4.5 <u>Examining stationarity of time series data</u>

Intuitively, a time series is stationary if the statistical properties (for example, the mean and the variance) of the time series are essentially constant through time.

If we have observed n values $y_1, y_2, ..., y_n$ of a time series, we can use a plot of these values (against time) to help us to determine whether the time series is stationary. If the n values seem to fluctuate with constant variation around a constant mean, then it is reasonable to believe that the time series is stationary. Otherwise, it is non-stationary.

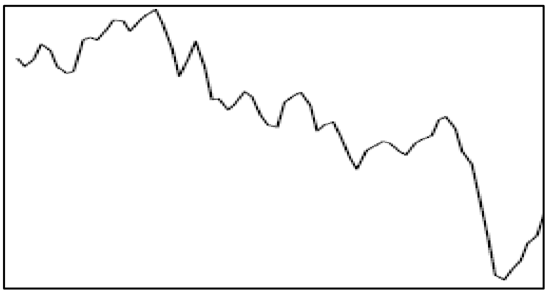
For example,

(a)



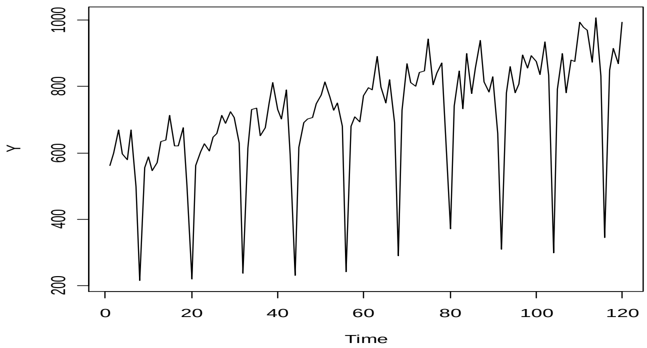
No evidence of a change in the mean over time, the series is stationary in the mean.

(b)



The mean of the series changes over time, hence non-stationary in mean.

(c)



The mean and variance of the series changes over time, hence non-stationary in both mean and variance.

Most of the probability theory of time series is concerned with stationary time series, and for this reason time-series analysis often requires one to transform a non-stationary series into a stationary so as to use this theory. For example, it may be of interest to remove the trend and seasonal variation from a set of data and then try to model the variation in the residuals by means of a stationary stochastic process.