1 Reference

- Applied Statistics by Gupta & Kapur
- Fundamental of Statistics Vol II Gun,Gupta Dasgupta

2 Introduction

A time series is a collection of observations made sequentially in time. Time series deal with statistical data which relate to successive interval or points of time. Time series can be represented hourly, daily, monthly, weekly, quarterly, yearly etc. As for example: monthly production of a particular commodity etc. Time series data usually refers economic data. It also applies in natural and social sciences. Thus, if the values of a phenomenon or variable at times $t_1, t_2, \cdots, t_n$ are $U_1, U_2, \cdots, U_n$, then the series

$$\begin{aligned} t &: t_1, t_2, t_3, \cdots, t_n \\ U_t &: U_1, U_2, U_3, \cdots, U_n \end{aligned}$$

constitute a time series. For example population of a country (U_t) in different years (time point) t.

3 Components of time Series

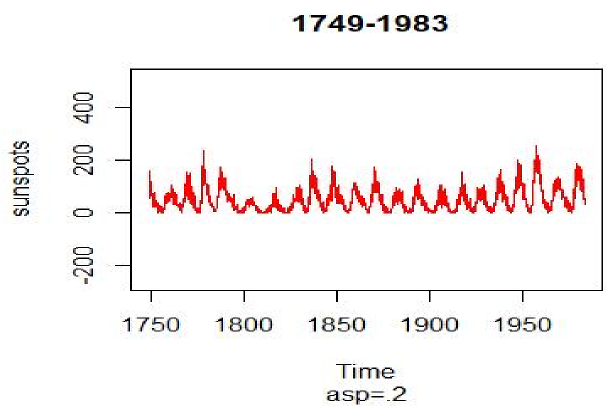


Figure 1: A time series plot

A graphical representation of time series reveals the change of variable over time (t).A series which exhibits no change during the time period under consideration will give just a horizontal line. But we are interested on those time series which are showing continual changes over time, giving an overall impression of haphazard movement (look at the plot). A critical study of time series contains a systematic progression as well as some haphazard movements.

In analytical language time series can be split in two primary part : 1. Systematic part 2. Unsystematic part or irregular part.

Systematic part is attributed to three different factors

- Secular Trend
- Seasonal Variation
- Cyclical Variation

In a given time series some or all of the above components may be present.

Separation of different time component is necessary because we may be interested in a particular component or we want to study the series after eliminating the effect of a particular component. Also, for predicting on future values of variable (forecasting) we need separation of time series component. Please remember, for forecasting time series only systematic part is needed.

In classical or traditional approach two types of format of model are used– Additive and Multiplicative.
Let us name Secular Trend:T_t, Seasonal Component: S_t, Cyclical Component:(C_t) and Irregular Component (I_t). Then
Multiplicative Time Series Model: $U_t = T_t \times S_t \times C_t \times I_t$
Additive Model: $U_t = T_t + S_t + C_t + I_t$.
Multiplicative decomposition is same as the additive decomposition of the logarithmic values of original time series. In practice, most of the series relating to economic data conform to the multiplicative model.

A time series model can be constructed in combined format also, like $U_t = T_tC_t + S_tR_t$ or $U_t = T_t + S_t + C_tR_t$. But this type of structures are not very common.

4 Uses of time series

The time series analysis is of great importance not only to economic study but also in various disciplines in natural, social and in physical sciences. There are many uses of time series analysis. Basically it can be applied to anything that changes over time. That includes:

- Financial data e.g. stock market indices
- Energy demand e.g. electricity demand
- Weather variables e.g. temperature, pressure etc.
- Epidemics
- Astronomical data
- Medical data like EEG
- Data collected by sensors like internet of things

Time series analysis helps studying the processes generating these data as well as forecast future values. It also helps us to compare the changes in the values of different phenomenon at different times or places etc.

5 Discussion On component in details:

5.1 Trend

By secular trend or simply trend we mean the general tendency of data to increase or decrease or constant during a long period of time. Generally in economy and business upward trend or downward trend is visible. But it should not be inferred that all of the time trend must show upward or downward tendency. Moreover, it is not necessary that the increasing or declining should be in the same direction through out the entire time period. Trend line might be changed. Trend usually measured the average pattern of a time series. For example, the study of trend in agriculture, industry, education, population of a country is quite necessary.

If the time series values cluster around a straight line then the average pattern or trend for that time series is linear trend–otherwise nonlinear or curved trend. In a linear trend, rate of growth or decline is constant.

The term–'long period of time' in defining trend is a relative form and does not have any specific choice how much long it is. In some cases, a period as small as a week is fairly long while in some other cases two/three years not enough to study the trend. Many a times trend is inspected just by visual guess.

5.2 Seasonal Component (variation)

Seasonal variation in a time series is due to the rhythmic forces which occurs periodically in the time series. For seasonal variation periodicity spans less than a year. Thus, seasonal variation in a time series can be recorded quarterly, monthly,weekly,daily even hourly. For example,traffic jam in a road crossing or sale of umbrellas during monsoon season are the common example of seasonal variation.

Generally, seasonal variation arises due to natural cause. On the contrary, sometimes man-made causes are the source of seasonal variation, like habits, fashions, customs and conventions of people in the society varying season to season. For instance, the sale of jewellery and ornaments goes up during the wedding season.

The main objective of measurement of seasonal variations is to isolate them from the trend and study their effects solely. As this seasonal variation repeats over time, separating it from the main series would enable us to focus on trend and other components of time series in better way. This removing of seasonal variation is called <u>Deseasonalization</u>. A time series where the seasonal component has been removed is called seasonal stationary. A time series with a clear seasonal component is referred to as non-stationary. Mathematically, this separation can be done by dividing the time series U_t by seasonal variations S_t.

$$T_t \times C_t \times I_t = \frac{U_t}{S_t}$$

5.3 Cyclical Variation

The oscillatory movement in a time series with period of oscillation **more than one year** are termed as cyclic fluctuation. One complete period is called a cycle. The cyclical movement in business cycle composed of boom (prosperity), regression (recession), depression, progression(recovery). This is called four phase business cycle. cyclical fluctuation, though more or less regular; are not certain always.

5.4 Irregular Component

Apart from the regular variations almost all the series contains another factor called random or irregular fluctuation. These fluctuations are purely random, unforeseen, unpredictable. The effect of earthquakes, tsunami,political upheavals, wars are considered as the random components in the population structure for a country.

6 Measurement of trend–discussion on methods

1. Graphical Method
2. Method of Semi averages
3. Method of curve fitting by least square method
4. Moving Average Method

6.1 Graphical Method

- The crudest way to draw a time series data
- free hand joining without any mathematical support

6.2 Method of of Semi Averages

In this method the whole data is divided into two parts with respect to time. In case of odd number of the time points we will omit the value corresponding to the middle year. Next we compute the arithmetic mean for each part and plot these two averages against the mid values of the respective period covered by each period. The line obtained on joining these two points is the required trend line and may be extended both ways to estimate intermediate or future values.

Method of semi averages is easy to understand and computationally simple. But this method assumes linear relationship between the plotted points–which may not exist. Moreover, the limitation of using arithmetic mean cause wrong trend estimation.

6.3 Curve fitting by least square method

To determine the trend line in terms of known functions, say like straight line,polynomial curve, exponential curve etc. the most common technique is least square principle. Least square principle is used for determining the constants of the functions. Apart from arithmetic scale sometimes semi logarithmic for T_t is also used.

6.3.1 Some known Curves

1. Linear Curve: $T_t = a + bt$
2. Parabolic Curve:$T_t = a + bt + ct^2$
3. Polynomial Curve:$T_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$
4. Exponential Curve: $T_t = ab^t$
5. Growth Curve:
   (a) Modified Exponential:$T_t = c + ab^t$
   (b) Gompertz Curve:$T_t = ab^{c^t}$
   (c) logistic curve: $T_t = \frac{k}{1+e^{a+bt}}$

All the constants a, b, c can be estimated through least square principle. In general for linear curve the least square estimates are extracted from the following normal equations.

$$\begin{aligned}
\sum_{t=1}^{k} T_t &= ka + b\sum_{t=1}^{k} t \\
\sum_{t=1}^{k} tT_t &= a\sum_{t=1}^{k} t + b\sum_{t=1}^{k} t^2
\end{aligned}$$

while for fitting k degree polynomial curve we have $(k+1)$ normal equations as follows.

$$\begin{aligned}
\sum_{t=1}^{k} T_t &= ka_0 + a_1 \sum t + a_2 \sum t^2 + \cdots + a_k \sum t^k \\
\sum_{t=1}^{k} tT_t &= a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 + \cdots + a_k \sum t^{k+1} \\
\cdots \sum t^k T_t &= a_0 \sum t^k + a_1 \sum t^{k+2} + a_2 \sum t^{k+3} + \cdots + a_k \sum t^2 k
\end{aligned}$$

For exponential fitting we will transform logarithmically , i, e, $Y_t = logT_t = loga + tlogb$, hence $Y_t = a' + b't$ where $a' = loga, b' = logb$.

Note 1. *How to decide what would be the required function of trend?*
Linear Trend *If U_t are monotonically increasing or decreasing with an absolute constant, then linear trend should be used.* $u_t - U_{t-1} = r \forall t$.
Exponential Trend *U_t's are increasing/decreasing by a constant proportion.* $\frac{U_t}{U_{t-1}} = r$.

Alternative Way

- If $\Delta U_t = constant$ linear trend
- If $\Delta^2 U_t = constant$ parabolic trend
- If $\Delta^k U_t = constant$ polynomial with degree k
- If $\Delta(logU_t) = constant$ exponential trend
- For modified exponential and Gompertz $\frac{\Delta U_t}{\Delta U_{t-1}}$ or $\Delta(logU_t)/\Delta(logU_{t-1})$ will be constant.

6.4 Merits of least squares

1. It is most popular and widely used method of fitting mathematical function.
2. Because of its mathematical character, these method completely eliminates any kind of judgemental bias.
3. Unlike moving average method, this method gives the estimated trend for any time period.

6.5 Growth Curve & their fitting

Various growth curves, modified exponential,Gompertz and logistic curve cannot be determined by the least square principle since all these cases, the number of parameters (constants to be determined) is more than the number of variables. So special technique has to be devised for these curves.

6.5.1 Modified exponential curve and its fitting

Modified exponential curve is given by $U_t = a + bc^t$ where a, b, c are the constants. We discuss one method of estimating the parameters.

6.6 Method of Partial Sum

The given time series data are split into three equal parts , each containing say n consecutive values of U_t, corresponding to $t = 1, 2, \cdots, n; t = n + 1, n + 2, \cdots, 2n$;and $t = 2n + 1, 2n + 2, \cdots, 3n$so that $S_1 = \sum_{t=1}^{n} U_t$, $S_2 = \sum_{t=n+1}^{2n} U_t$ and $S_3 = \sum_{t=2n+1}^{3n} U_t$. Substituting for U_t by the equation of the curve, we get

$$\begin{aligned} S_1 &= \sum_{t=1}^{n} U_t \\ &= \sum_{t=1}^{n} (a + bc^t) \\ &= na + bc\frac{c^n - 1}{c - 1} \qquad (1) \end{aligned}$$

Similarly,

$$\begin{aligned} S_2 &= \sum_{t=n+1}^{2n} U_t \\ &= \sum_{t=n+1}^{2n} (a + bc^t) \\ &= na + bc^{n+1}\frac{c^n - 1}{c - 1} \qquad (2) \end{aligned}$$

$$
\begin{aligned}
S_1 &= \sum_{t=2n+1}^{3n} U_t \\
&= \sum_{t=2n+1}^{3n} (a+bc^t) \\
&= na + bc^{2n+1}\frac{c^n-1}{c-1} \qquad (3)
\end{aligned}
$$

Subtracting (1) from (2) and (2) from (3)we get respectively

$$S_2 - S_1 = bc\frac{(c^n-1)^2}{c-1} \qquad (4)$$

and

$$S_2 - S_1 = bc^{n+1}\frac{(c^n-1)^2}{c-1} \qquad (5)$$

Dividing (5) by (4) we have

$$
\begin{aligned}
\frac{S_3-S_2}{S_2-S_1} &= c^n \\
c &= (\frac{S_3-S_2}{S_2-S_1})^{\frac{1}{n}}
\end{aligned}
$$

Substituting c^n in (5), we get $S_2-S_1 = \frac{bc}{c-1}\left[\frac{S_3-S_2}{S_2-S_1}-1\right]^2 = \frac{bc}{c-1}\left[\frac{S_3-2S_2+S_1}{S_2-S_1}\right]^2$ which leads to

$$b = \frac{c-1}{c}\frac{(S_2-S_1)^2}{(S_3-2S_2+S_1)}.$$

Finally, substituting the value of b and c in (1)we get $a = 1/n[S_1 - bc\frac{c^n-1}{c-1}] = \frac{1}{n}\left[\frac{S_1S_3-S_2^2}{S_3-2S_2+S_1}\right]$(Show every step of simplification).

6.7 Fitting of Gompertz Curve

The equation of Gompertz curve is $U_t = ab^{c^t}$. Taking logarithm in both sides in the above equation we have

$$
\begin{aligned}
logU_t &= loga + ctlogb \\
Y_t &= A + BC^t \qquad (6)
\end{aligned}
$$

where $Y_t = logU_t, A = loga, B = logb$. Now eq. (6) is a modified exponential curve and constants A, B and C can be estimated by the method of partial sum. Finally the constants of Gompertz curve can be given by

$$a = e^A, b = e^B, c = \left(\frac{S_3-S_2}{S_2-S_1}\right)^{\frac{1}{n}}$$

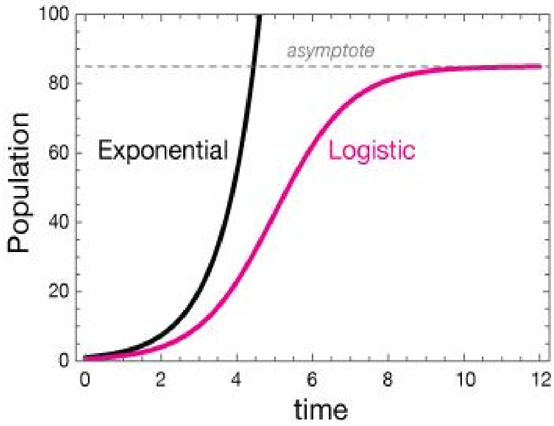


Figure 2: Logistic Curve

6.8 Fitting of Logistic Curve

Logistic curve, most widely used growth curve, looks like an elongated "S". A symmetric logistic curve is given by

$$U_t = \frac{k}{1 + e^{(a+bt)}}, b < 0$$

where a, b, k are constants.

Estimation of the parameters can be done by method of three selected points.

6.9 Method of Three Selected Points

The given time series data is first plotted on a graph paper and a trend line is first drawn by the freehand method. Three ordinates U_1, U_2, U_3 are now taken from the trend line corresponding to selected equidistant points of time, say $t = t_1, t = t_2$ and $t = t_3$ respectively such that $t_2 - t_1 = t_3 - t_2$.

Substituting the values $t = t_1, t_2$ and t_3 in the formula of logistic curve, we get respectively $U_1 = \frac{k}{1+e^{a+bt_1}}, U_2 = \frac{k}{1+e^{a+bt_2}}, U_3 = \frac{k}{1+e^{a+bt_3}}$ which leads

$$\begin{aligned} a + bt_1 &= log(\frac{k}{U_1} - 1) \\ a + bt_2 &= log(\frac{k}{U_2} - 1) \\ a + bt_3 &= log(\frac{k}{U_3} - 1) \end{aligned} \quad (7)$$

$\Rightarrow$

$$b(t_2 - t_1) = log\left[\frac{(k/U_2) - 1}{(k/U_1) - 1}\right]$$
$$b(t_3 - t_2) = log\left[\frac{(k/U_3) - 1}{(k/U_2) - 1}\right] \tag{8}$$

Since the points are equidistant,i,e,$t_2 - t_1 = t_3 - t_2$, we get

$$log\left[\frac{(k/U_2) - 1}{(k/U_1) - 1}\right] = log\left[\frac{(k/U_3) - 1}{(k/U_2) - 1}\right]$$
$$(\frac{k}{U_3} - 1)(\frac{k}{U_1} - 1) = (\frac{k}{U_2} - 1)^2$$
$$U_2^2(k - U_3)(k - U_1) = U_1U_3(k - U_2)^2$$
$$k^2(U_2^2 - U_1U_3) = k(U_2^2(U_1 + U_3) - 2U_1U_2U_3)$$

Since $k \neq 0$, we get $k = \frac{U_2^2(U_1+U_3)-2U_1U_2U_3}{U_2^2-U_1U_3}$.From (7) and (8) we get respectively

$$b = \frac{1}{t_2 - t_1} log\left[\frac{(k - U_2)U_1}{(k - U_1)U_2}\right]$$
$$a = log\left(\frac{k - U_1}{U_1}\right) - bt_1$$

.

6.10 Moving Average Method

Moving average method consists in measurement of trend by smoothing out the fluctuations of the data by means of a moving average. Moving average of extent (period)m is a series of successive averages (arithmetic means) of m terms at a time , starting with 1st, 2nd, 3rd term etc. Thus the first average is the mean of the first m terms, the second is the mean of the m terms from 2nd to $(m + 1)$th term, the third is the mean of the m terms from 3rd to $(m + 2)$ th term and so on.

If m is odd=$(2k+1)$, say, moving average is placed against the mid-value of the time interval it covers, i.e. against $t = k + 1$ and if m is even, say =$2k$(say), it is placed between two middle values of the time interval,i,e, between $t = k$ and $t = k+1$. In the latter case the moving average does not coincide with an original time period as it sits in between two time points. So in order to centering , a further 2-point moving average is executed.The graph obtained on plotting the moving average against time gives trend.

Although moving average method is quite flexible and meaningful as it smooths out the observed time series at each averaging point it suffers from a few drawback.

6.10.1 Drawbacks

1. It does not provide trend values for all the terms. For a MA of extent $2k+1$, first k terms and the last k terms will be unestimated.

2. It can't be used for forecasting or prediction of future trend.

3. It fails if the underlying trend is nonlinear (Moving average is suitable when trend is linear or approximately linear).

4. If the underline oscillatory movement is haphazard MA can not perform well.

6.10.2 How to fix the order of MA :a thumb rule

In MA method the main problem lies in determining the period of the MA. It has been established mathematically that if the fluctuations are regular and periodic, then the moving average completely eliminates the oscillatory movement provided the extent of MA is exactly equal to the period of oscillation or a multiple of it. So the technique is to find the peak of the time series and calculating the distance between two successive peaks which gives periodicity of cycle. Finally average out all such obtained periods of all cycles. That average will be the suitable period of moving average.

6.11 Deseasonalization/ Elimination of Seasonality Component

We have a raw time series U_t. After trend elimination, next important task is to eliminate the seasonal fluctuations (period$<$1 yr).

Why deseasonalisation is needed?

1. To isolate seasonal variation means to get an idea of seasonal variation means to get an idea of seasonal fluctuation over the original time

series. So that, one can figure out the actual value of the observation when there is no seasonality.

2. For some practical, commercial purpose deseasonalisation is needed also. For example, for a business, knowing the seasonal pattern of sales gives the idea of forecasting for future time periods, so that a businessman can maintain his proper inventory.

Note 2. ***Seasonal Index*** *The degree of seasonal variation for a particular season is called seasonal index of that season. This measures how a particular season behave as compared with the average seasonal effect.*

Remark For a quarterly data, total seasonal index is theoretically 400 considering that each season contributes equal effect (100). On the other hand, for yearly data, theoretically total seasonal index is 1200.
There are four methods of deseasonalization.

- Method of simple average
- Ratio to moving average method
- Ratio to trend method
- Link Relatives method

6.11.1 Method of Simple Average

1. Average the data by years and months
2. Compute the averages $\overline{x_i}$ (for yearly data $i = 1(1)12$ and for quarterly $i = 1(1)4$)
3. Compute the general average: For monthly data it is $\overline{x} = \frac{1}{12}\sum_{i=1}^{12}\overline{x_i}$.
4. seasonal index for ith season is $\frac{\overline{x_i}}{\overline{\overline{x}}} \times 100$.

Note 3. *The assumption for using this process is that the raw data is free of cyclical component.*

6.12 Ratio to trend method

This method is an improvement over the simple averages method. This is based on the assumption that seasonal variation for any given month is constant factor of the trend. The measurement of seasonal variation by this method consists in the following steps.

1. Obtain the trend values by the least square method by fitting a mathematical curve, straight line or 2nd degree polynomial etc.

2. Express the original data as the percentage of the trend values. Assuming the multiplicative model, these percentages will contain the seasonal, cyclic and irregular components.

3. The cyclic and irregular components are then wiped out by averaging the percentages for different months (quarters) if the data are monthly (quarterly), thus leaving us with indices of seasonal variations. Either arithmetic mean or median can be used for averaging.

4. Finally these indices are adjusted to a total 1200 for monthly or 400 for quarterly data by multiplying them throughout by a constant k given by $k = \frac{1200}{Total of the indices}$ and $k = \frac{400}{total of indices}$.

If the series has little or less cyclical fluctuations this method of deseasonalization works well. On the contrary, if there exist pronounced cyclic fluctuations, the trend value obtained through least square method will not be the correct one which follows biases ratio to trend method.

6.13 Ratio to moving averages

Moving average eliminates periodic movements if the extent (period) of moving average is same as the period of oscillatory movement. Thus for a monthly data, a 12 point moving average should completely eliminate the seasonal movement if they are of constant pattern and intensity. The method of getting seasonal indices by moving average involves the following steps:

1. calculate the centered 12 month moving average of the data. These moving average values will give estimates of the combined effects of trend and cyclic variations.

2. Express the original data (except for 6 months in the beginning and 6 months at the end) as percentages of the centered moving average uses. Using multiplicative model, these percentages would then represent the seasonal and irregular components.

3. The preliminary seasonal indices are now obtained by irregular or random component by averaging these percentages. Arithmetic mean can be used.

4. The sum of these indices will not be, in general, 1200. So an adjustment is done to make the sum of the indices 1200 by multiplying throughout by a constant factor=1200/sum of all seasonal variations, i.e. by expressing the preliminary seasonal indices as the percentage of their arithmetic mean.

6.13.1 Merits and Demerits

Of all of the methods of measuring seasonal variations, ratio to moving average method is the most satisfactory, flexible and widely used method. These indices do not fluctuate so much as the indices by the ratio to trend method.

This method does not completely utilise the data, e.g. for monthly data, first and last six months do not have moving average seasonal indices.

6.14 Link Relatives Methods of Deseasonalization

Link relatives in seasonality detection method are the relative increment or depletion of seasonality on one season to immediately next season.

6.14.1 Method

1. To calculate the link relative for each of the season. What is link relatives?

$$Link\,relative\,for\,any\,month = \frac{Current\,month's\,figure}{Previous\,month's\,figure} \times 100.$$

At the first step translate the original data into link relatives(L.R.)

2. Average the link relatives for each month(quarter) if the data are monthly(quarterly).

3. Convert those average link relatives (A.L.R.)into chain relatives on the base of the first season. Chain relative (C.R.) for any season is obtained as follows (considering C.R. of Jan=100)

$$C.R.\,for\,February = \frac{A.L.R.\,of\,Feb \times C.R.\,of\,Jan}{100} = L.R.\,of\,Feb$$

$$C.R.\,for\,March = \frac{A.L.R.\,of March \times C.R.\,of\,Feb}{100}$$

$$C.R.\,for\,Dec = \frac{A.L.R.\,of\,Dec \times C.R.\,of\,Nov}{100}.$$

4. Now , by taking this December value as a base, a new chain relative for January has to be found.

$$C.R.\,for\,Jan = \frac{A.L.R.\,of\,Jan \times C.R.\,of\,Dec}{100}.$$

   Obviously this new chain relative for January will not be 100 (the previous chain relative of Jan).

5. Correction Factor $d = \frac{1}{12}[New\,C.R.\,for\,January - 100]$. New chain relative. Then assuming linear trend, the correction factor for February, March,..., Dec is $d, 2d, \cdots, 11d$ respectively, subtract them accordingly. (in quarterly data it will be $d, 2d, 3d$ and $4d$.)

6. Hence we find corrected C.R. for each month. These are the corrected seasonal indices. Check if the sum of corrected seasonal indices is 1200 (in quarterly data 400), if not then proceed for adjustment factor

7. Adjustment factor= 1200/total of seasonal indices. Adjusted Seasonal indices=corrected seasonal indices$\times$ Adjustment factor.

6.14.2 Merits

This methods utilises data more completely than moving average method. Though not as simple as the moving average method, the actual method of link relatives are much less extensive.

7 Measurement of Cyclic Component

7.1 Crude Method

1. Any time series U_t can be written as $U_t = T_t \times S_t \times C_t \times I_t$. Estimate T_t by moving average method.

2. Also estimate the seasonal component preferably by 12-point (4-point moving average method.

3. Divide U_t by $T_t \times S_t$.

4. The resulting value gives C_t, the cyclic component.

The use of moving average method of suitable period will eliminate out the random component I_t. Choosing the period of moving average in estimating T_t is difficult one, so this method is seldom used.

7.2 Harmonic Analysis (Spectral Analysis)

Harmonic analysis provides a sophisticated method of determining the cyclic component of a time series. In harmonic analysis a time series is imagined as a continuous wave. Any wave function can be represented by Fourier series, i.e., a series of sums of sine and cosine functions. Thus we decompose a stationary time series (U_t) with period of oscillation λ into sinusoids (by Fourier expansion) as follows

$$U_t = a_0 + a_1 sin\frac{2\pi t}{\lambda} + a_2 sin\frac{2\pi}{\lambda}.2t + a_3 sin\frac{2\pi}{\lambda}.3t + \cdots + b_1 cos\frac{2\pi t}{\lambda} + b_2 cos\frac{2\pi}{\lambda}.2t + \cdots$$

where the constants are determined by Euler-Fourier equations

$$\begin{aligned} a_i &= \frac{1}{\lambda}\int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} U_t\, sin\left(\frac{2\pi}{\lambda}it\right) dt, \text{where } i = 1, 2, \cdots \\ b_i &= \frac{1}{\lambda}\int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} U_t\, cos\left(\frac{2\pi}{\lambda}it\right) dt, \text{where } i = 0, 1, 2, \cdots \end{aligned}$$

For equally spaced sample of size n (i.e. $t = 1, 2, ..n$), these integrations can be replaced by finite sum where

$$a_i = \frac{2}{n}\sum_{t=1}^{n} U_t\, sin\left(\frac{2\pi}{\lambda}it\right)\ (i = 1, 2, \cdots)$$

$$b_i = \frac{2}{n}\sum_{t=1}^{n} U_t \, cos\left(\frac{2\pi}{\lambda}.it\right) \; (i = 1, 2, \cdots)$$

$$a_0 = \frac{1}{n}\sum_{t=1}^{n} U_t.$$

For instance, if the period of oscillation is 12 months and $U_1, U_2, \cdots, U_12$ is the series or average of series for a number of years, then the constants a_i's and b_i's are given by

$$a_0 = \frac{1}{12}\sum_{t=1}^{12} U_t$$

$$a_i = \frac{1}{12}\sum_{t=1}^{12} U_t \, sin\left(\frac{2\pi}{12}it\right), \; (i = 1, 2, \cdots, 6)(why?)$$

$$b_i = \frac{1}{12}\sum_{t=1}^{12} U_t cos\left(\frac{2\pi}{12}it\right), \; (i = 1, 2, \cdots, 5)$$

So far λ(periodicity) is regarded as a known constant. What would be the dominant periods in a time series?

To address this question **Periodogram** analysis provides an elegant method of determining optimum choice of λ.

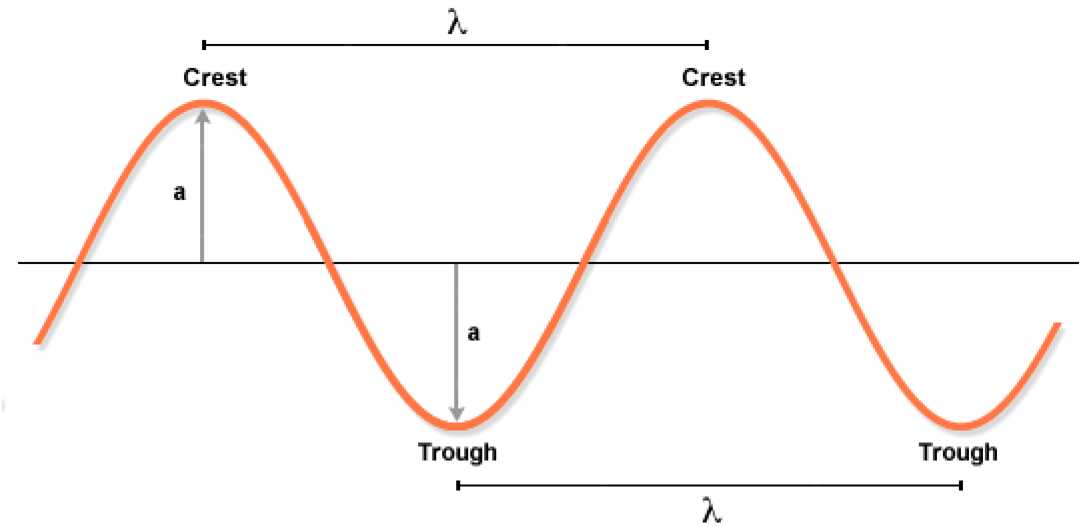
7.3 Periodogram Analysis

Any time series can be expressed as a combination of cosine and sine waves with **differing periods** (how long it takes to complete a full cycle) and amplitudes (maximum/minimum value during the cycle). This fact can be utilized to examine the periodic (cyclical) behavior in a time series.

Definition 1. ***Periodogram:*** *A periodogram is used to identify the dominant periods (or frequencies) of a time series. This can be a helpful tool for identifying the dominant cyclical behavior in a series, particularly when the cycles are not related to the commonly encountered monthly or quarterly seasonality.*

Definition 2. ***Period:*** *(T) is the number of time periods required to complete a single cycle of the cosine (sine) function.*

Definition 3. ***Frequency:*** *Frequency is $w = 1/T$. It is the fraction of the complete cycle that's completed in a single time period.*



Let us consider a time series U_t in which the trend and the seasonal component have been eliminated.Imagine fitting a single sine wave to a time series U_t, observed in discrete time. U_t has two components, one periodic with periodλ and the random component ϵ_t. So

$$U_t = asin\frac{2\pi}{\lambda}t + \epsilon_t \tag{9}$$

Also, we consider some conditions.

1. homoscedastic i.i.d. sequence of random components

2. $Cov(\epsilon_t, sin\frac{2\pi\, t}{\lambda}) = 0$; $Cov(\epsilon_t, cos\frac{2\pi\, t}{\lambda}) = 0$.

Let us consider the coefficients

$$\begin{aligned} A &= \frac{2}{n}\sum_{t=1}^{n} U_t cos\frac{2\pi t}{\mu} \\ B &= \frac{2}{n}\sum_{t=1}^{n} U_t sin\frac{2\pi t}{\mu} \end{aligned} \tag{10}$$

where μ is arbitrary and define $S^2(\mu) = A^2 + B^2$ which is known as **intensity** corresponding to the trial period μ. Substituting from (??) in (1) and using the conditions (stated above)

$$\begin{aligned}
A &= \frac{2}{n}\sum_{t=1}^{n}(a\,sin\frac{2\pi t}{\lambda}+\epsilon_t)cos\frac{2\pi t}{\mu} \\
&= \frac{a}{n}\sum_{t=1}^{n}2sin\frac{2\pi t}{\lambda}cos\frac{2\pi t}{\mu} \text{ as } Cov(\epsilon_t, cos\frac{2\pi\, t}{\lambda})=0 \\
&= \frac{a}{n}\sum_{t=1}^{n}2sin\alpha\; t\,cos\beta\, t \text{ where } \alpha=\frac{2\pi}{\lambda} \text{ and} \beta=\frac{2\pi}{\mu}
\end{aligned}$$

So $A=\frac{a}{n}\sum_{t=1}^{n}[sin(\alpha+\beta)t+sin(\alpha-\beta)t]$. Let $S=\sum_{t=1}^{n}sin(\alpha+\beta)t$. On further computation we get

$$\begin{aligned}
S.sin\left(\frac{\alpha+\beta}{2}\right) &= \frac{1}{2}\sum_{i=1}^{n}2sin(\alpha+\beta)t\,.sin\left(\frac{\alpha+\beta}{2}\right) \\
&= \frac{1}{2}\sum_{t=1}^{n}\left[cos\{(\alpha+\beta)t-\frac{\alpha+\beta}{2}\}-cos\{(\alpha+\beta)t+\frac{\alpha+\beta}{2}\}\right] \\
&= \frac{1}{2}[\{cos\left(\frac{\alpha+\beta}{2}\right)-cos\frac{3(\alpha+\beta)}{2}\}+\{cos\frac{3(\alpha+\beta)}{2}-cos\frac{5(\alpha+\beta)}{2}\} \\
&+ \cdots+\{cos\frac{(2n-1)(\alpha+\beta)}{2}-cos\frac{(2n+1)(\alpha+\beta)}{2}\}] \\
&= \frac{1}{2}\left[cos\frac{\alpha+\beta}{2}-cos\frac{(2n+1)(\alpha+\beta)}{2}\right] \\
&= sin\{\frac{(n+1)(\alpha+\beta)}{2}\}.sin\frac{n(\alpha+\beta)}{2}
\end{aligned}$$

Thus $S=\frac{sin\frac{n(\alpha+\beta)}{2}.sin\frac{(n+1)(\alpha+\beta)}{2}}{sin(\frac{\alpha+\beta}{2})}$. Similarly we can get the value of $\sum_{t=1}^{n}sin(\alpha-\beta)t$. Substituting in the expression of A we get

$$A=\frac{a}{n}\left[\frac{sin\frac{n(\alpha+\beta)}{2}sin\frac{(n+1)(\alpha+\beta)}{2}}{sin\frac{\alpha+\beta}{2}}+\frac{sin\frac{n(\alpha-\beta)}{2}sin\frac{(n+1)(\alpha-\beta)}{2}}{sin\frac{\alpha-\beta}{2}}\right]$$

If $\alpha\neq\beta$, then $A\longrightarrow 0$ as $n\to\infty$. However if $\alpha-\beta\to 0$, then for large n, we get

$$A=lim_{n\to\infty}\frac{a}{n}.\text{[some finite quantity]}+lim_{n\to\infty}\frac{a}{n}\frac{sin\frac{n(\alpha-\beta)}{2}sin\frac{(n+1)(\alpha-\beta)}{2}}{sin\frac{\alpha-\beta}{2}}$$

which equals $a\,sin\frac{(n+1)(\alpha-\beta)}{2}$ for large n, (since $lim_{\theta\to 0}\frac{sin\,n\theta}{sin\theta}=n$). Thus for large n,

if $\alpha \neq \beta$, $\lambda \neq \mu$ then $A \to 0$ and
if $\alpha \to \beta$,$\lambda \to \mu$, then $A \to a\, sin \frac{(n+1)(\alpha-\beta)}{2}$.
Similarly, it can be shown that for large n
if $\alpha \neq \beta$, $\lambda \neq \mu$ then $B \to 0$ and
if $\alpha \to \beta$,$\lambda \to \mu$, then $B \to a\, cos \frac{(n+1)(\alpha-\beta)}{2}$.

Thus if the arbitrary period μ is exactly the period of oscillation (λ) of the series, then $S^2(\mu) = a^2$ or $S(\mu) = a$. The intensity remains small if the experimental period μ is away of true period λ.

7.3.1 Drawing Periodogram

From given time series $U_1, U_2, \cdots, U_n$ calculate A and B for different μ from 0 to n and compute $S(\mu)$. The graph obtained on plotting $S(\mu)$ against μ is known as periodogram. Maximum values of $S(\mu)$ would provide the true period of the time series. The obvious drawback of periodogram analysis is its huge calculations.

7.4 A Simple Example by R

A plot of $S^2(\mu)$, as spikes, against $1/\mu$ (frequency) is a Fourier line spectrum. The raw periodogram in R is obtained by joining the tips of the spikes in the Fourier line spectrum to give a continuous plot and scaling it so that the area equals the variance.

7.4.1 R code

```
t <- seq(0,100,by=0.5)
 U_t <- cos(2*pi*t/16) + 0.75*sin(2*pi*t/5)
#The variable U_t is made up of two underlying periodicities: the first at a frequen
 #16 (one observation completes 1/16'th of a full cycle, and a full cycle is complet
 #and the second at a frequency of 1/5 (or period of 5)
plot(t,U_t,'l')}
 spectrum(U_t)
#the spectrum function goes further and automatically removes a linear trend from th
#calculating the periodogram
```

Question

1. Extract "'Sunspots"' data in R-studio.
2. What type of data it is?
3. Convert it in a monthly time series data.
4. Present the data from 1750 to 1900.
5. Plot it starting from the year 1750 ending at the year 1900.
6. Report five newest observations
7. Report the observations of the year 1800.
8. Install the package "zoo". Find 12 point centered moving averages and plot with the original series.
9. Plot periodogram(spectrum).
10. What is the period? Give an estimate.

Question In Acadly find the .csv file on Natural gas prices in USA. Notice it is a monthly data.

1. Plot it (you may convert the times as time point by r code ¿ ts(data set))
2. Calculate moving average of order 12.
3. Plot original data and moving average points on same graph paper.
4. Plot periodogram(spectrum).
5. What is the period?

For the above problem when extracting the csv file and plotting out, two plots will be showing off as R system take month as a variable as well. To discard month from the first plot, use the following codes.

```
y<-read.csv("path of the file", header=TRUE, stringAsfacstor=False)
# This stringAsfactor consider that date information are characters.
y.ts<-ts(y$Price(the variable on which the plot will be made))
plot(y.ts)
```