1 Introduction

A time series is a sequence of observations, which are ordered in time (or space). In electrical engineering literature, time series are called signals. There are two types of time series data:

- **Continuous**, wherein observations are made at every instant of time, for example electrocardiograms, lie detectors, etc. A continuous time series can be denoted as \(x(t)\). Continuous time series are sampled (measured) at discrete time points and are ultimately treated as discrete series.

- **Discrete**, wherein observations are made at (usually) equispaced intervals. A discrete time series can be represented as \(x_t : t = 1, 2, ..., N\) in which the subscript \(t\) indicates the time at which the datum \(x_t\) was observed.

Time series data provide useful information about the physical, biological, social or economic systems generating the time series, such as:

- **Economics**: share prices, profits, imports, exports, stock exchange indices,
- **Sociology**: school enrollments, unemployment, crime rate,
- **Environment**: amount of pollutants, such as suspended particulate matter (SPM), sulphur dioxide, nitric oxide, etc. in the environment,
- **Meteorology**: rainfall, temperature, wind speed,
- **Epidemiology**: number of SARS cases over time,
- **Medicine**: blood pressure measurements over time for evaluating drugs to control hypertension,
- **Physical sciences**: sun spots,
- **Hydrology**: water flows,

and many more.
Here are a few examples of plots of time series data:

![Fig. 1: Time series 1](image1)

![Fig. 2: Time series 2](image2)

### 1.1 Types of time series

There are two types of time series: univariate and multivariate. Univariate time series are those where only one variable is measured over time, whereas multiple time series are those, where more than one variable are measured simultaneously.

### 1.2 Components of time series

Typically, a time series comprises four components:

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

- **Seasonal component** - Seasonal fluctuations of known periodicity. It is the component of variation in a time series which is dependent on the time of the year. It describes any regular fluctuations with a period of less than one year. For example, the costs of various types of fruits and vegetables, and average daily rainfall, all show marked seasonal variation.

- **Cyclic component** - Cyclical variations of non-seasonal nature, whose periodicity is unknown.

- **Irregular component** - Random or chaotic noisy residuals left over when other components of the series (trend, seasonal and cyclical) have been accounted for.

Trend and seasonality, though conceptually distinct, are essentially entangled. The value of the series at time $t$ essentially depends on its value at time $t-1$, with the result that trend and periodic components are inextricably mixed up. Hence, it is not possible to isolate one without trying to isolate the other.


1.3 Objectives of time series analysis

An observed time series can be assumed as the realization of a stochastic process. Once we understand how the process operates, we can develop a mathematical model to predict the future values of the time series. Thus, there are two main objectives of time series analysis:

1. To understand the underlying structure of the time series by breaking it down to its components,
2. To fit a mathematical model and then proceed to forecast the future.

Basically, there are two approaches to time series analysis, which are associated with the time domain (i.e. trend component) or the frequency domain (i.e. periodic component).

The time domain approach represents time series as a function of time. Its main concern is to explore whether the time series has a trend (rising or declining) and if so, to fit a forecasting model.

The frequency domain approach is based on the assumption that the most regular, and hence predictable, behavior of a time series is likely to be periodic. Thus, the main concern of this approach is to determine the periodic components embedded in the time series.

The choice between the frequency domain and the time domain depends essentially upon the types of questions that are being asked in different fields of study. For example, economists have relatively greater interest in the time domain, whereas communication engineers have greater interest in the frequency domain. However, combining these two approaches would yield a better understanding of the data.

2 Time Domain Analysis

2.1 Definitions and concepts

2.1.1 Stationarity

Stationarity is a critical assumption in time series models. Stationarity implies homogeneity in the sense that the series behaves in a similar way regardless of time, which means that its statistical properties do not change over time. More precisely, stationarity implies that the joint probability distribution of the process is invariant over time. However, in practice, a much weaker definition of stationarity, often referred to as weak stationarity is used.

The condition of weak stationarity requires that the elements of the time series should have a common finite expected value and that the autocovariance of two elements should depend only on their temporal separation. Mathematically, these conditions are:

\[
\text{Mean} = \mu \quad \text{and} \quad \text{Variance} = \sigma^2 \quad \text{are constant for all values of } t
\]

\[
\text{Covariance} \; \gamma(x_s, x_r) \; \text{is a function of } (s - r) \; \text{only.} 
\]  

(1)

Hereafter, we would drop the adjective “weak” and a stationary process would always mean a weak stationary process.

Remark: The term weak stationarity is also called second order stationarity, wide sense stationarity or covariance stationarity.
Stationarity of a time series can be judged by just looking at its time plot. Intuitively, a time series would be stationary if its time plot appears similar at different points along the time axis, but there are other statistical and graphical means to evaluate the stationarity.

2.1.2 Random time series

A time series, in which the observations fluctuate around a constant mean, have a constant variance and are statistically independent, is called a random time series. In other words, the time series does not exhibit any pattern:

- The observations do not trend upwards or downwards,
- The variance does not increase or decrease over time,
- The observations do not tend to be larger in some periods than in other periods.

A random model can be written as

\[ x_t = \mu + e_t \] (2)

where

- \( \mu \) is a constant, the average \( x_t \)
- \( e_t \) is the residual (or error) term which is assumed to have a zero mean, a constant variance, and \( e_t \)'s are statistically independent.

One can examine whether a time series is random or not by the following procedures:

1. Visually, whether the time series plot shows any trend or not.
2. Visually by looking at the Correlogram of the time series (explained in § 2.1.3).
3. Statistically, testing whether the observed series could have been generated by a random stochastic process, for example, statistical test based on the counts of “Turning Points”.

Turning Points

Count the number of peaks or troughs in the time series plot. A peak is a value greater than its two neighbours. Similarly, a trough is a value less than its two neighbours. The two (peak and trough) together are known as Turning Points.

Consider the time series \( [x_1, x_2, x_3, ..., x_N] \). The initial value \( x_1 \) cannot define a turning point since we do not know \( x_0 \). Similarly, the final value \( x_N \) cannot define a turning point since we do not know \( x_{N+1} \). Three consecutive values are required to define a turning point. In a random series, these three values can occur in any of six possible orders and in any four of them, there would be a turning point. The probability of finding a turning point in any set of three values is thus

\[ E(p) = E(C_i) = \frac{2(N-2)}{3} \]

Now, define a counting variable \( C_i \)

\[ C_i = \begin{cases} 
1 & \text{if } x_i < x_{i+1} > x_{i+2} \text{ or } x_i > x_{i+1} < x_{i+2} \\
0 & \text{otherwise.} 
\end{cases} \] (3)
The number of turning points \( p \) in the series is given by

\[
p = \sum_{i=1}^{N-2} C_i
\]

\[
E(p) = E(C_i) = \frac{2(N - 2)}{3}
\]

where \( E \) is the expectation operator.

If the observed number of Turning Points is more than the expected value, \( \frac{2(N - 2)}{3} \), then they could not have arisen by chance alone. In other words, the series is not random. In order to test whether the difference between the observed and expected number of turning points is statistically significant, we have to calculate the variance of \( p \). From combinatorial algebra,

\[
Var(p) = \frac{16N - 29}{90}
\]

\[
Std(p) = \sqrt{\frac{16N - 29}{90}}
\]

We can test an observed value against the expected value from the following statistic, which has standard Normal distribution

\[
Z = \frac{p - E(p)}{Std(p)} = N(0, 1)
\] (4)

### 2.1.3 Autocorrelation

The most distinguishing feature of time series is that successive values are not independent; they are correlated with one another, i.e. they are serially correlated or autocorrelated Autocovariance and autocorrelation functions are important tools for describing the serial (or temporal) dependence structure of a univariate time series.

Autocorrelation is a measure of the dependence of time series values at a certain time on the values at another time. It is the Pearson correlation between all the pairs of points in the time series with a given separation in time or lag. Positively autocorrelated series are sometimes referred to as persistent because high values tend to follow high values and low values tend to follow low values. Negatively autocorrelated series are characterized by reversals from high to low values from one time segment to the next, and vice versa.

Autocorrelation functions of deterministic data (like sine wave) persist over all time displacements, whereas autocorrelation functions of stochastic processes tend to zero for large time displacement (for zero-mean time series).

The first order correlation (i.e. lag = 1) is the correlation coefficient of the first \( N - 1 \) observations \( [x_t : t = 1, 2, ..., N - 1] \) and the next \( N - 1 \) observations \( [x_t : t = 2, 3, ..., N] \). It is given by the following formula:

\[
r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}(1))(x_{t+1} - \bar{x}(2))}{\sqrt{\sum_{t=1}^{N-1} (x_t - \bar{x}(1))^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}(2))^2}}
\] (5)

where \( \bar{x}(1) \) is the mean of the first \( N - 1 \) observations and \( \bar{x}(2) \) is the mean of the last \( N - 1 \) observations.
For reasonably large $N$, the difference between $\bar{x}_1$ and $\bar{x}_2$ would be negligible, and $r_1$ can be approximated by

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

(6)

where $\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t$ is the overall mean.

Equation (6) can be generalized for computing the correlation between observations separated by $k$ time units

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

(7)

The array of autocorrelation coefficients $r_k$ provides crucial information about the internal structure of the time series. The plot with $k$ as abscissa and $r_k$ as ordinate is called a Correlogram.

The standard error of autocorrelation for lag $k$ is computed by the following formula:

$$SE_k = \sqrt{\frac{1}{N} \sum_{i=1}^{k} r_i^2}$$

(8)

The value of the autocorrelation function at lag 0 is always equal to 1. The horizontal band about zero represents the approximate 95% confidence limits for $H_0 : \rho = 0$. If no autocorrelation estimate falls outside the strip defined by the two dotted lines, and the data contain no outliers, one may safely assume that there is no serial correlation.

The autocorrelation plot issued by WinIDAMS shows a band ($\pm$2 standard error) around the mean zero, indicating 95% confidence limits.

![Autocorrelation plot](image)

**Fig. 3: Autocorrelation plot**

**Interpretation of correlogram**

Correlogram summarizes characteristic features of the time series, viz. randomness, rising or declining trend, oscillation, etc. Below are some keys for interpreting the correlogram.
• **Random series** - If a time series is completely random, then for large \( N \), \( r_k = 0 \) for all non-zero values of \( k \).

• **Short term correlation** - Stationary time series exhibit short term correlations, characterized by a fairly large value of \( r_k \) followed by two or three more coefficients, which while statistically significant tend to become successively smaller. Values of \( r_k \) tend to approximate zero for large \( k \).

• **Alternating series** - If successive observations of a time series tend to alternate on different sides of the overall mean, the correlogram would also tend to oscillate.

• **Nonstationary time series** - If a time series has a trend, then the values of \( r_k \) would not decrease to zero, except for large values of \( k \). However, it would be desirable to detrend the series, otherwise its characteristics would be swamped by the trend.

• **Seasonal fluctuations** - If a time series is characterized by seasonal fluctuations, then the correlogram would also exhibit oscillations at the same frequency.

### 2.1.4 Autocovariance

In practice, autocorrelation coefficients are computed from autocovariance coefficients \( \gamma_k \). Autocovariance function of a time series at lag \( k \) is defined by the following equation:

\[
\gamma_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \overline{x})(x_{t+k} - \overline{x})
\]

Autocorrelation coefficient \( r_k \) at lag \( k \) is given by

\[
r_k = \frac{\gamma_k}{\gamma_0}
\]  

### 2.1.5 Cross-correlation

Cross-correlation is a measure of similarity between two time series. Mathematically speaking, it is the linear correlation coefficient between two time series as a function of time lag between the two series. Consider \( N \) pairs of observations on two time series \( x_t \) and \( y_t \).

Cross-covariance function is given by the following formulae:

\[
\gamma_{xy}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \overline{x})(y_{t+k} - \overline{y}) \quad \text{for} \quad k = 1, 2, 3, \ldots, N - 1
\]

\[
\gamma_{yx}(k) = \frac{1}{N} \sum_{t=1-k}^{N} (x_t - \overline{x})(y_{t+k} - \overline{y}) \quad \text{for} \quad k = -1, -2, -3, \ldots, -(N - 1)
\]
where

\[ \gamma_{xy}(k) \] is the cross-covariance when \( y_t \) lags \( x_t \)
\[ \gamma_{yx}(k) \] is the cross-covariance when \( x_t \) lags \( y_t \)

\( N \) is the series length

\( \bar{x} \) and \( \bar{y} \) are the sample means

\( k \) is the lag.

Cross-correlation is the cross-covariance scaled by the variances of the two series

\[ r_{xy}(k) = \frac{\gamma_{xy}(k)}{\sqrt{\gamma_{xx}(0)\gamma_{yy}(0)}} \] (12)

where \( \gamma_{xx}(0) \) and \( \gamma_{yy}(0) \) are the sample variances of series \( x_t \) and \( y_t \).

While the autocovariance function and autocorrelation function are symmetric functions (value at lag \( k \) equals value at lag \( -k \)), the cross-covariance function and cross-correlation function are asymmetric functions (value at lag \( k \) is not equal to that at \( -k \)). The asymmetry necessitates the two parts of Equation (11). The first part applies to \( y_t \) lagging \( x_t \), and the second part to \( x_t \) lagging \( y_t \).

### 2.1.6 White Noise

The most fundamental concept in time series analysis is “White Noise”. A stationary time series for which \( x_t \) and \( x_{t+k} \) are uncorrelated is called White Noise. Such a process will have autocorrelation function \( r_k \)

\[ r_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \] (13)

### 2.2 Pre-processing of time series data

In order to fit a time series model to the data, it is often essential to first transform the data to render them “well-behaved”. By this we mean that the transformed data can be modeled by a zero-mean, stationary process. We can usually decide whether a particular time series is stationary just by looking at its time plot. Intuitively, a time series “looks” stationary if the time plot of the series appears “similar” at different points along the time axis. Look at the plot of the time series (Airline passengers’ data).
It can be easily seen that not only the number of passengers tends to increase (a positive trend), but also the spread or variance tends to increase over time. Hence, it would be necessary to remove the trend (i.e. detrend the time series) and stabilize its variance.

### 2.2.1 Detrending

The detrending or removal of the trend from the time series is a statistical or mathematical operation, which involves the following steps:

- Estimate the trend by fitting linear or nonlinear regression models,
- Subtract the time series generated by the trend from the original time series.

Stabilization of variance can be achieved by transformation or smoothing of time series.

The following alternatives are also available for detrending:

- First differencing,
- Trend identification and detrending.

### 2.2.2 First differencing

A time series that is non-stationary in mean can be made stationary by taking the first-difference. The first-difference is simply the difference of the value of the series at times \( t \) and \( t - 1 \),

\[
y_t = x_t - x_{t-1}
\]

where \( x_t \) is the original time series and \( y_t \) is the first-differenced series. The number of observations in the differenced series \( y_t \) would be one less than the number of observations in the original series. But, if the series is also not stationary in the rate of change of the mean (i.e. slope), stationarity can be achieved by taking the second difference, or the first difference of the first difference

\[
U_t = y_t - y_{t-1}
\]
It should, however, be borne in mind that each successive differencing will decrease the variance of the series, but at some point, higher-order differencing will have an opposite effect. When variance increases, it means that the series has been over-differenced.

Usually a lag of 1 is used to produce the differenced series. A lag of 12 can be used to seasonally adjust monthly data. Similarly, a lag of 4 can be used to seasonally adjust quarterly data.

![Fig. 5: Plot of time series Temperature](image1.png)

![Fig. 6: Autocorrelation plot of time series Temperature after smoothing (Difference method)](image2.png)

It can be easily seen that all the points are within the 95% confidence band.

### 2.2.3 Trend identification and detrending

If a time series is monotonic (consistently increasing or decreasing) and does not contain considerable error (i.e. random noise), the trend can be modeled as a function of time. However, if the time series contains considerable random noise then it is necessary to filter the noise prior to trend identification.

The simplest and most widely used filter is the linear filter, which has been found to approxi-
mate adequately many monotonic time series. However, if there is a clear monotonic nonlinear component, other functional forms should be considered.

After fitting the trend equation, the question still remains, how to remove the trend from the time series? There are two options:

1. Subtract the value of the trend from the original data, which yields a time series of residuals from the trend. This “difference” option is attractive for simplicity, and for giving a convenient breakdown of the variance. The residual series is in the same units as the original series, and the total sum of squares of the original data can be expressed as the trend sum-of-squares and residual sum-of-squares.

2. The other option is to compute the ratio of the original time series to the trend equation. The “ratio” option is attractive for some kinds of data since it tends to remove 'trend in variance' that might accompany the 'trend in mean'. Another advantage is that the ratio is dimensionless. An important disadvantage is that ratio-detrending is only feasible for nonnegative series. Another disadvantage is that ratio-detrended series may explode to high value if the trend line approaches close to zero.

2.2.4 Filtering

A filter is a system that separates specific components of a time series (for example, random noise or specific range of frequencies). The moving average is the most common filter for smoothing the time series, mainly because it is the easiest filter to understand and use. In spite of its simplicity, the moving average filter is optimal for reducing random noise, while retaining the trend, which makes it the most important filter for time domain analysis. However, the moving average is the worst filter for frequency domain analysis, with little ability to separate one band of frequencies from another.

As the name implies, the moving average filter operates by averaging a number of points from the input signal (i.e. a given time series) to produce each point in the output signal (filtered series). In equation form, this is written as

\[ y_t = \sum_{k=0}^{W} b_k x_{t-k} \] (14)

where \( x_t \) is the input series, \( y_t \) is the filtered series, \( b_k \) are known as the filter coefficients, and \( W \) is width of the filter.

An unweighted moving average with a relatively small number of elements (say three to five) will have its smoothing effect without destroying the seasonality present in a series. A moving average with a larger number of elements or with weights designed to emphasize the elements toward the center of the subset would possibly be even more efficient in removing the irregular variation, but would also tend to destroy any seasonality.

To sum up, stationarity can be induced by the following sequence of steps:

1. Remove the trend by first differencing or by subtracting the trend curve.

2. The detrended series might still fluctuate with increasing amplitude. In that case, successive differencing would not induce stationarity. In such a case, stationarity may be induced by transforming the detrended series into logarithms.
3 Measurement of Trend

The following methods are available for studying the trend:

1. Visual inspection of the time series plot,
2. Curve fitting by least squares method (i.e. Parametric regression models),
3. Moving Averages.

3.1 Visual inspection

Look at the following plots. Figure 7 shows a rising trend, whereas Figure 8 does not show any trend.

Fig. 7: Increasing trend

Fig. 8: No trend
3.2 Parametric regression models

Linear trend in a monotonic time series can be estimated by fitting a simple regression model

\[ x_t = At + B + e_t \]  \hspace{1cm} (15)

where

- \( x_t \) is the value at time \( t \)
- \( B \) is a constant
- \( A \) is the regression coefficient
- \( e_t \) are the regression residuals.

The parameters \( A \) and \( B \) can be estimated using the least squares method described in Chapter 4 (Guide to Advanced Data Analysis). Estimates of parameters \( A \) and \( B \) are given by the following equations:

\[
\hat{A} = \frac{\sum_{t=1}^{N} (t - \bar{t})(x_t - \bar{x})}{\sum_{t=1}^{N} (t - \bar{t})} \hspace{1cm} (16)
\]

\[
\hat{B} = x_0 - \hat{A}\bar{t}
\]

where

- \( \bar{t} = \frac{1}{N} \sum_{t=1}^{N} t = \frac{N + 1}{2} \)
- \( \bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t \)

\( A \) is the slope of the regression line.

The magnitude of \( A \) indicates the trend (or average rate of change) and its sign indicates the direction of the trend. \( B \) indicates the value at time 0.

It should be noted that \( e_t \)'s are not independent; they are correlated. This violates a basic assumption of regression analysis. Nonetheless, parameter estimates are unbiased, but probability estimates are not unbiased. Hence, the statistical significance of the parameters cannot be tested. Appropriately, the Trend option in the Time Series module of WinIDAMS reports only the parameter estimates, but not their statistical significance.

A disadvantage of the linear regression model is that it cannot detect trends that are nonlinear, but still monotonic. Hence, one has to consider quadratic or other nonlinear functional forms.
Time Series module of WinIDAMS incorporates six regression models:

1. Linear
   \[ x_t = At + B \]

2. Quadratic
   \[ x_t = At^2 + Bt + C \]

3. Exponential
   \[ x_t = \text{Exp}(At + B) \]

4. Logarithmic
   \[ x_t = A \log_n (t + B) \]

5. Logistic
   \[ x_t = \frac{A}{1 + \text{Exp}(B - Ct)} \]

6. Power
   \[ x_t = A(t + B)^C \]

Models (2) - (6) can be fitted through least squares procedure after suitable transformation of the data.

Diagnostic checking

After fitting a parametric regression model, it would be desirable to detrend the series by abstracting the regression model and testing for stationarity to check the adequacy of the fitted model.

3.3 Autoregressive model

Autoregressive model of a time series is based on the assumption that each value of the series depends only on a weighted sum of the previous values of the same series plus “noise”. Notice that, in this model, time is a sequencing variable, and not a predictor variable as in parametric regression models described in § 3.2.

\[ x_t = \delta + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} + e_t \tag{17} \]

where

\[ \delta = (1 - \sum_{i=1}^{p} \varphi_i) \mu \]

\( x_t \) is the original time series, or its nonlinear transform (for example, logarithmic or square root); the series is assumed to be stationary with mean \( \mu \)

\( e_t \) is Gaussian white noise

\( \varphi_i \) are autoregression coefficients

\( p \) is the order of the autoregression model; usually, \( p \) is much less than the length of the series \( (p \ll N) \).
Thus, an autoregressive model is simply a linear regression of the current value of the series on one or more prior values of the same series. Several options are available for analyzing autoregressive models, including standard linear least squares techniques. They also have a straightforward interpretation.

The time series $x$ is called an autoregressive process of order $p$ and is denoted as AR($p$) process. Notice that AR($p$) is a $p$-th order Markov process.

All values of autoregression coefficients $\varphi_i$ do not result in a stationary process. For example, AR(1) process with $\varphi = 1$ is a random walk, which is known to be nonstationary. It can be easily seen that AR(1) process would be stationary only if $|\varphi| < 1$.

For an AR($p$) process, the condition for stationarity is that the complex roots of the equation

$$\psi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p$$

must lie outside the unit circle.

The most common method for deriving the coefficients involves multiplying the definition above by $x_{t-k}$, taking the expectation values and normalizing, which yields a set of linear equations, called the Yule-Walker equations relating the parameters of an AR model with the sequence of autocorrelation coefficients. In the matrix form, Yule-Walker equations can be written as

$$
\begin{bmatrix}
1 & r_1 & r_2 & \cdots & r_{p-1} \\
r_1 & 1 & r_2 & \cdots & r_{p-2} \\
r_2 & r_1 & 1 & \cdots & r_{p-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
\varphi_1 \\
\varphi_2 \\
\varphi_3 \\
\vdots \\
\varphi_p
\end{bmatrix}
= 
\begin{bmatrix}
r_1 \\
r_2 \\
r_3 \\
\vdots \\
r_p
\end{bmatrix}
$$

where $r_k$ is the autocorrelation coefficient at lag $k$. Note that the diagonal=1.

The matrix of coefficients in Equation (19) is a Toeplitz matrix; that is, the elements on each diagonal are all the same. Because of this property, there is a recursive method that allows one to obtain estimates for a $k$-th order model from the estimates of the $(k-1)$-th model in a fast and accurate manner. The method is called the Levinson or Levinson-Durbin algorithm.

Selecting the order of the model

In practice, the order of the autoregressive model is not known, and there is no straightforward method to determine the appropriate order. Hence, it would be desirable to solve the Yule-Walker equations for several values of $p$ ranging from $p = 1$ through $p_{\text{max}}$, where $p_{\text{max}}$ is sometime 10 or more. As one increases the order of the model, the root mean square (RMS) error generally decreases sharply up to a certain value and then more slowly. A value just after the point at which RMS error flattens out is usually an appropriate order. (Similar to the Scree plot in factor analysis). There are of course more formal methods for choosing the model order. The most common is Akaike Information coefficient (AIC).

4 Frequency Domain Analysis

4.1 Introduction

So far, we have studied stationary time series in terms of quantities that are functions of time - covariance and correlation functions. This approach is called time series analysis in the time domain.
Another approach is to analyze the time series in Fourier space or in the frequency domain. The techniques used in the frequency domain fall under the general rubric of spectral analysis and the basic mathematical tool is the Fourier transform. Squared modulus of the Fourier transform is called the spectrum.

Spectral analysis essentially decomposes a stationary time series \([x_t]\) into a sum of sinusoidal components of different frequencies with uncorrelated random coefficients. In conjunction with this decomposition, there is a corresponding decomposition into sinusoids of the autocovariance function (acvf). However, instead of presupposing the existence of a finite set of known frequencies, spectral analysis allows representation of all possible frequencies.

The spectrum contains no new information beyond that contained in the autocovariance function (acvf). The spectrum can be computed mathematically by the transformation of acvf but the spectrum and acvf present information on the variance of the time series from complementary viewpoints. The acvf summarizes information in the time domain and the spectrum in the frequency domain.

The spectrum of a time series is not only an important theoretical concept but it is also an important practical tool to gain insight into the periodicities of the data. To apply spectral analysis, it is essential that the time series, \(x_t\), satisfies the following theoretical conditions:

1. Time series is covariance stationary, i.e. \(E(x_t) = \mu\) and \(cov(x_t, x_{t-k}) = \gamma_k\). Both \(\mu\) and \(\gamma_k\) must exist and be independent of \(t\).

2. Multiple estimates of \(\mu\) and \(\gamma_k\) must converge as the length of the sampled series increases. This requirement is called Ergodicity condition.

**Objectives of spectral analysis**

- To detect unknown hidden frequencies in the periodic time series,
- To provide useful descriptive statistics,
- As a diagnostic tool to indicate which further analyses might be relevant,
- To check postulated theoretical models.

**4.2 Notation and terminology**

While discussing the frequency domain, it would be useful to start with definitions pertaining to waves. Consider the simple example of a cosine wave (Fig. 9).
The peaks are high points in the wave, and troughs are the low points. The time from one peak (trough) to another peak (trough) is called time period, noted as $T$. The vertical distance from zero to the peak is called the amplitude, noted as $R$. The variance is proportional to the square of the amplitude (i.e. Variance=$R^2/2$). The phase is the offset (in time) of the peaks or troughs from a fixed point in time. The angular frequency $\omega$ describes the number of cycles per unit time, measured in radians, where $2\pi$ radians corresponds to a complete cycle of the wave (peak to peak).

In practical applications, frequency is denoted as $f$, the number of cycles per unit time. The two frequency measures are related as follows:

$$f = \frac{\omega}{2\pi}$$

The wavelength, or period, of the cosine wave is the distance from peak to peak ($\lambda$), and is the reciprocal of the frequency

$$\lambda = \frac{1}{f}$$

For a discrete process measured at unit intervals of time, there is no loss of generality in restricting $\omega$ to the range $(0, \pi)$, since

$$\cos(\omega t + k\pi t) = \cos(\omega t) \quad k, t \text{ integers with } k \text{ even}$$

$$\cos(\omega t + k\pi t) = \cos(\pi - \omega t) \quad k, t \text{ integers with } k \text{ odd}$$

Hence, variations at frequencies higher than $\pi$ cannot be distinguished from the corresponding frequency in $(0, \pi)$. The frequency $\omega = \pi$ is called Nyquist frequency. For a discrete process measured at equal intervals of time length, $\Delta t$, the Nyquist frequency is $\pi/\Delta t$.

### 4.3 Mathematical tools

#### 4.3.1 Fourier Transform (FT)

Fourier transform (FT) is a mathematical function that can be used for mapping a time series from the time domain into the frequency domain. In effect, FT decomposes a waveform or a function into sinusoids of different frequencies which sum to the original waveform. It identifies or distinguishes different frequency sinusoids and their respective amplitudes.

Fourier transform can be expressed as $x(t) \rightarrow X(f)$ according to

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \, dt$$

Equation (20) defines the Fourier transform of $x(t)$ and Equation (21) defines the inverse Fourier transform that recovers $x(t)$ back from $X(f)$. These equations are at the heart of spectral analysis and they are so tightly connected that they are usually called the Fourier transform pair, symbolically represented as
\[ x(t) \leftrightarrow X(f), \text{ that is } x(t) \text{ corresponds to } X(f) \]

\( X(f) \) and \( x(t) \) are respectively called the frequency domain and time domain representations of the signal. It is customary to use a lower case symbol for the time domain function and an upper case symbol for the corresponding function of frequency. The terms in the integrands exclusive of \( x(t) \) and \( X(f) \) are called kernels of the transforms. In the Fourier transform pair, the kernels differ only slightly. The sign of the exponent in the Fourier transform is negative and that in the inverse transform is positive.

If frequency is measured in radians per unit time (angular frequency \( \omega \)) instead of cycles per unit time, Equations (20) and (21) can be written as

\[
X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \tag{22}
\]

\[
x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega \tag{23}
\]

### 4.3.2 Discrete Time Fourier Transform (DTFT)

If we only know the function \( x(t) \) at evenly spaced points \( x(nT) \) where \( T = 1/f_s \) and \( f_s \) is the sampling rate, the Fourier transform would be given by

\[
X(f) = \sum_{n=-\infty}^{\infty} x_n e^{-i2\pi f_n T} \tag{24}
\]

This function is periodic with period \( f_s \). The sampled function can be recovered by inverse transform

\[
x_n = \frac{1}{f_s} \int_{f_s}^{f_s} X(f) e^{i2\pi f_n T} df \tag{25}
\]

### 4.3.3 Discrete Fourier Transform (DFT)

If we have a finite sequence \([x_n]\) of sample from a function \( x(t) \), the discrete Fourier transform is defined by

\[
X(f_k) = \sum_{n=0}^{N-1} x(t_n) e^{-i2\pi nk/N} \tag{26}
\]

where \( N \) is the length of sequence \([x_n]\) and \( k \) is the \( k \)-th frequency sample.

The inverse transform is

\[
x(t_n) = \sum_{k=0}^{N-1} X(f_k) e^{i2\pi nk/N} \tag{27}
\]

Although these functions are described here as complex series, real valued series can be represented by setting the imaginary part equal to 0. In general, the transform into the frequency domain will be a complex valued function with magnitude and phase

\[
\text{Magnitude} = \| X(f_k) \| = (X_{\text{real}} \times X_{\text{real}} + X_{\text{imag}} \times X_{\text{imag}})^{0.5}
\]
Phase \( = \tan^{-1}\left(\frac{X_{\text{mag}}}{X_{\text{real}}}\right) \)

The magnitude measures how strongly the oscillation at frequency \( \omega \) is represented in the data. The strength of the periodic component is more often represented by the periodogram defined further in this chapter.

### 4.3.4 Fast Fourier Transform (FFT)

Fast Fourier transform (FFT) is an efficient implementation of the discrete Fourier transform (DFT) for highly composite transform lengths. When \( N \) is a power of 2, computational complexity drops from \( O(N)^2 \) for DFT to \( O(N \log_2 N) \) for FFT. This means a dramatic savings in the computational effort for large \( N \). If the length of the time series is not a power of 2, the length can be artificially increased by adding zeros up to the next power of 2. This is called zero-padding. However, if the time series is small (say, a few thousand cases), then do not worry; the analysis will typically take only a few seconds anyway.

### 4.3.5 Power

#### Continuous signal

The total power of a signal is the same whether we compute it in the time domain or in the frequency domain.

\[
\text{Total power} = \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |X(\omega)|^2 \, d\omega
\]

(29)

In order to compute the power contained in the interval between \( \omega \) and \( \omega + d\omega \), we do not distinguish between positive and negative frequency, but rather regard \( \omega \) as varying between 0 and \( +\infty \). One sided Power Spectral Density is given by

\[
p(\omega) = |X(\omega)|^2 + |X(-\omega)|^2
\]

(30)

When the function \( x(t) \) is real, then the two terms in Equation (30) are equal, so that

\[
p(\omega) = 2|X(\omega)|^2
\]

(31)

Discrete analogues of power are defined below.

#### Discrete time series

Power of a time series is defined as the average of time-series energy. If \( x_k \) is the \( k \)-th value of a time series of \( N \) samples with sampling period \( \Delta t \), its energy \( E \) is defined as

\[
E = \sum_{k=0}^{N-1} |x_k|^2 \Delta t
\]

(32)

When the series \( x_k \) is produced by a stationary stochastic process of infinite duration rather than by a deterministic waveform of finite duration, the energy is usually infinite, and the quantity of interest is the power \( p \) defined as the energy per unit of time

\[
p = \frac{E}{N \Delta t} = \frac{1}{N} \sum_{k=0}^{N-1} |x_k|^2
\]

(33)

For zero-mean time series, the power is equal to the variance of the sample of the \( N \) values \( x_k \).
4.4 Spectral distribution function

According to the Wiener-Khinchin theorem, for any stationary stochastic process with autocovariance function $\gamma(k)$, there exists a monotonically increasing function $F(\omega)$, such that

$$\gamma(k) = \int_{0}^{\pi} \cos(\omega k) dF(\omega) \quad (34)$$

Equation (34) is called spectral representation of the autocovariance function and $F(\omega)$ is called the spectral distribution function.

$F(\omega)$ has a direct physical interpretation. $F(\omega) = \text{contribution to the variance of the series which is accounted for by frequencies in the range } [0, \omega]$. $F(\omega)$ increases monotonically between $\omega = 0$ and $\omega = \pi$, and hence is similar to a cumulative distribution function (cdf). Scaling $F(\omega)$ by the variance $\sigma_{x}^{2}$, we get the normalized spectral distribution function $F^{*}(\omega)$ defined as

$$F^{*}(\omega) = \frac{F(\omega)}{\sigma_{x}^{2}} \quad (35)$$

which gives the proportion of the variance accounted for by the frequencies in the range $[0, \omega]$ and reaches a maximum of 1.0, since $F^{*}(\pi) = 1$.

4.5 Spectral density function or spectrum

By differentiating the spectral distribution function we get the (power) spectral density function, which gives the variance associated with each frequency

$$f(\omega) = \frac{dF(\omega)}{d\omega} \quad (36)$$

The term “spectral density function” is often shortened to spectrum. The adjective “power” is often omitted. “Power” comes from the application of spectral analysis in electrical engineering.

4.6 Estimation of spectrum

At this point, it is essential to mention that spectral analysis is basically concerned with the

- Estimation of the unknown periodicities of the time series from the data,
- Quantification of the relative importance of different frequency bands to the variance of the process.

It should be recognized that the spectrum being estimated is not really the spectrum of the observed series, but the spectrum of the unknown infinitely long series from which the observed series have been sampled at discrete time points.

Various methods for estimation of spectrum can be categorized as follows:

- Nonparametric methods,
- Parametric methods,
- Subspace methods.
Nonparametric methods are those in which the estimate of the power spectral density (PSD) is made directly from the sequence of observations itself. The simplest such method is the periodogram. An improved version of the periodogram is the Welch’s method.

Parametric methods are those in which the signal is assumed to be the output of a linear system driven by white noise. Examples are the Yule-Walker autoregressive (AR) method and the Burg method. These methods estimate the PSD by first estimating the parameters (coefficients) of the linear system that hypothetically “generates” the time series.

Subspace methods, also known as high-resolution methods or super-resolution methods, generate frequency component estimates for a signal based on an eigenanalysis or eigendecomposition of the correlation matrix.

The most popular algorithm is based on the direct computation of the squared modulus of the Fourier transform of the time series (often termed as periodogram estimator). This algorithm is implemented in the Time Series module of WinIDAMS.

4.6.1 Blackman-Tukey method

The estimation of spectrum on the basis of autocovariance of the observed data is not consistent, meaning that the variance of the estimate does not decrease as the sample size $N$ increases, no matter how long the series is. Another problem is that the autocovariance at high lags is uncertain, and the method does not discount the higher lags. As a result, the (estimated) spectrum fluctuates wildly from one frequency to another and, therefore, it becomes extremely difficult to interpret the spectrum.

The Blackman-Tukey method circumvents this problem by applying the Fourier transform to a truncated, smoothed $acvf$ rather than to the entire $acvf$. The method consists of taking a Fourier transform of the truncated sample $acvf$ using a weighting procedure, which gives less weight to the values of the $acvf$ at high lags. Such an estimator is given by

$$
\hat{S}(\omega) = \frac{1}{\pi} \left( \gamma_0 + \frac{2}{M} \sum_{k=1}^{M} w_k \gamma_k \cos(\omega k) \right)
$$

(37)

where

- $\gamma_k$ is a sequence of autocovariance functions
- $w_k$ is a set of weights called the lag window
- $M(< N)$ is called the truncation point.

From Equation (37) we see that $acvf$ estimates at lower lags are weighted by a weighting function $w_k$, which gives decreasing weights to higher lags, such that the higher lag $acvf$ values are discounted. Several window functions have been proposed in the literature, including the Tukey window. A crucial question is how to choose $M$. The bias, variance and bandwidth depend upon the truncation point $M$. The selection criteria are inherently conflicting and one has to strike a balance between resolution on the one hand and bias and variance of the spectral estimates on the other.

- **Bias** - Bias refers to the tendency of spectral estimates to be less extreme (both highs and lows) than the true spectrum, which means that peaks are not as high as they should be, and troughs not as low as they should be. Thus, the bias is reflected in the flattening of the estimated spectrum. It should be borne in mind that: smaller $M \Rightarrow$ greater bias.
• **Variance** - Smaller the truncation point $M$, smaller is the variance of the spectral estimates.

• **Bandwidth** - Roughly speaking, bandwidth is the width of the spectral window, which limits the frequency resolution of the spectrum. Resolution means the ability of the spectrum to distinguish between two spectral features that can be identified. Spectral estimates apply to a band of frequency, and the width of that band is the bandwidth. Thus if the bandwidth is wide, peaks nearby with similar frequencies might not be resolvable. The bandwidth is a function of truncation point $M$. Smaller $M$ implies greater bandwidth, i.e. decreased resolution.

### 4.6.2 Smoothed periodogram method

This method circumvents the Fourier transformation of the autocovariance function by direct Fourier transformation of the observed time series. This method is more popular than the method based on Fourier transform of autocovariance function. Squaring the magnitude of the results of Fast Fourier transform yields the estimate of the periodogram

$$\hat{S}(\omega_k) = \frac{1}{N} \left( \sum_{n=0}^{N-1} x(t_n) e^{-i\omega_k t_n} \right)^2$$

where

- $t_n$ is the $n$-th sampling point
- $N =$ number of frequency samples
- $\omega_k$ is the $k$-th frequency sample measured in radians per unit time.

The periodogram provides an approximation to the power spectrum. In this respect, the procedure breaks down the signal into sine wave components across a frequency range. The residuals of the least squares fit to the data are then summed. The period is chosen by minimization of this sum. The periodogram is a wildly fluctuating estimate of the spectrum with high variance. For a stable estimate, the periodogram must be smoothed.

### Limitations of periodogram

The estimation of power spectral density (PSD) as outlined above is computationally efficient and generally yields reasonable results. But, there are certain inherent performance limitations of the FFT approach.

• **Variance** - An inherent problem of periodogram estimates is that the variance is large, of the order of PSD squared. Moreover, the variance does not decrease as $N \to \infty$. There are two approaches to decrease the variance of periodogram estimates:

  1. Averaging in the time domain,
  2. Averaging in the frequency domain.

Averaging in the time domain is called Welch method, which involves breaking the signal into segments and then averaging the periodograms of these sections. Averaging in the frequency domain is called band averaging.
- **Spectral leakage** - The spreading of power across the spectrum is known as "leakage", i.e. artificially high power estimates at frequencies away from the true peak frequencies. FFT uses an implicit window (boxcar or rectangular window) function, which is equal to one over the period of interest and zero elsewhere. The boxcar function turns on and off quite rapidly. If we estimate this function as a sum of sines and cosines, we find that the estimate needs to have significant higher frequency components to deal with the jumps between 0 and 1. These higher frequency components cause contamination of the original signal.

This problem can be solved by using a suitable window that can preserve maximum power at the centre of the bin and almost zero power at the extremities of the bin. WinIDAMS uses Welch window, which is a parabolic function. (Also See § 4.7)

**Smoothing of periodogram**

- **Smoothness** - The raw (unsmoothed) periodogram is a rough estimate of the spectrum. It is *non-consistent* in the sense that the power estimate at the individual frequency fluctuates with \( N \), making difficult its interpretation. To build a spectral estimator which is more stable we turn to the technique of windowing among others. This technique is employed to smoothen all abrupt variations and to minimize the spurious fluctuations generated every time a series is truncated. The result of windowing is the *smoothed spectrum*. Thus, the smoothed spectrum at \( k \) is nothing but the periodogram seen through a window opened on a convenient interval around \( k \).

The raw periodogram is of little utility because of the high variance of the spectral estimates. Smoothing the periodogram with filters of various spans results in a spectrum much smoother in appearance than the raw periodogram. However, excessive smoothing obscures important spectral details; insufficient smoothing would leaves erratic and unimportant details in the spectrum. (Also See § 4.7)

- **Stability** - The stability of the spectral estimate is the extent to which estimates computed from different segments of a series agree, or, in other words, the extent to which irrelevant fine structure in the periodogram is eliminated. High stability corresponds to low variance of the estimate, and is attained by averaging over many periodogram ordinates.

- **Spectral resolution** - It is the ability of the spectrum to represent the fine structure of the frequency properties of the series. The fine structure is the variation in the spectrum between closely spaced frequencies. For example, narrow peaks are part of the fine structure of the spectrum.

Smoothing the periodogram with a filter averages over adjacent periodogram estimates, and consequently lowers the resolution of the spectrum. The wider the filter, greater is the smoothing and greater is the decrease in resolution. If two periodic components in the series were close to the same frequency, the smoothed spectrum might be incapable of identifying, or *resolving*, the individual peaks.

### 4.7 Windowing

It is but obvious that in real cases, where we cannot observe \( x(t) \) on the interval \((-\infty, \infty)\), we are looking at the function through a “window”. The window can be represented by a function \( w_t \). The window function for the ideal analytic case where we know the function on the interval \((-\infty, \infty)\), and the more realistic case where we know the function only on the interval \((0, T)\), are shown below.
Infinite data

In the case of infinite data, the window function is 1 for all $t$ and we get an exact representation of $x(t)$.

Window function: $w(t)x(t) = 1 \times x(t) \Leftrightarrow S(\omega)$

![Fig. 10: Infinite data](image)

Finite sample of data

In this case we view the time series $x(t)$ through a finite window, and therefore we do not get an exact measure of the spectrum, but something that is affected by the geometry of the window we are looking through.

![Finite sample of data](image)

In that case we do not obtain the true spectrum but a convolution of the true spectrum with a weighting window. This acts as smoothing on the true spectrum. The degree of smoothing depends on the length of the data window $T$. Shorter the length of the window, greater is the smoothing, but that would be at a certain cost. The side lobes of the frequency window would lead to spectral leakage.

In the case of rectangular window, the response function does peak strongly at the central frequency, but significant negative side lobes also occur. This means that spectral analysis will introduce spurious oscillations at higher and lower frequencies which are out of phase with the actual oscillation.

Ideally, a frequency window should have the following characteristics:

- A high concentration in the central lobe, which requires a large $T$,
- Small or insignificant side lobes, i.e. a smooth time window without sharp corners.

A rectangular window leaves the time series undistorted, but can seriously distort the frequency spectrum. A tapered window distorts the time series but may yield a more representative frequency spectrum.

A number of window functions are proposed in the literature. WinIDAMS’ module *Time Series Analysis* incorporates Welch window, which is a parabolic window (which falls to zero at the edges and is maximum at the middle).

**Welch window** looks like the following figure:

![Welch window](image)

Welch window is defined by the following function:

\[
W(n) = 1 - \left( \frac{n - N/2}{N/2} \right)^2
\]  

(39)

### 4.7.1 Welch Averaging method

Suppose \( P(f) \) is the spectral density of time series \( [x_t : t = 1, 2, ..., N] \). Let us take possibly overlapping segments of length \( L \) with starting points \( D \) units apart. Let \( k \) denote the number of segments chosen, such that

\[
(k - 1)D + L = N
\]

The segments are defined as follows for \( j = 0, 1, 2, ..., L - 1 \):

**First segment**

\[ x_1(j) = x(j) \]

**Second segment**

\[ x_2(j) = x(j + D) \]

**k-th segment**

\[ x_k(j) = x(j + (k - 1)D) \]
Now, select a data window $W(j)$ for $j = 0, 1, 2, ..., L - 1$ and form the sequences

$$x_1(j)W(j), ..., x_k(j)W(j)$$

Take the Fourier transform of these sequences and compute $k$ modified periodograms

$$S_k(f_n) = \frac{L}{U} |x_k(n)|^2$$

where

$$f_n = \frac{n}{L} \quad n = 0, 1, 2, ..., L/2$$

and

$$U = \sum_{j=0}^{L-1} W^2(j)$$

The spectral estimate is the average of these periodograms.

Averaging of periodograms tends to decrease the variance of the estimate relative to a single periodogram estimate of the entire data record. Generally, the variance is inversely proportional to the number of segments whose periodograms are being averaged. Although, overlap between segments tends to introduce redundant information, this effect is diminished by the use of a nonrectangular window, which reduces the importance or weight given to the end samples of segments (i.e. the samples that overlap).

### 4.7.2 Maximum Entropy method

The determination of spectral density (Fourier transform of autocorrelation) requires a complete knowledge of the covariance structure, but in most practical situations, we have only a finite length of observations of the time series, upon which to estimate the spectral density. A number of procedures have been proposed for estimating spectral density, but the method proposed by Burg, known as Maximum Entropy Method (MEM) has received the greatest attention.

MEM is based on the extrapolation of a segment of known autocorrelation function for lags which are not known. Suppose that autocorrelations for lags up to $k$ are known, then the question arises as to how autocorrelations for lags greater than $k$ should be specified so that the entire sequence of autocorrelations is positive semidefinite. Burg argued that the extrapolation should be made in such a way that the extrapolated autocorrelation function has the maximum entropy. Consequently, the time series (whose autocorrelations up to $k$ lags are known) would be the most random. In other words, the power spectral density would be the flattest of all spectra.

MEM imposes fewest constraints on the unknown time series by maximizing its randomness, thereby resulting in minimum bias. MEM has also been extended to the case of periodogram estimation through direct Fourier transform of the data sequence.

Caveat: The MEM is very efficient for detecting frequency lines in stationary time series, but if the time series is not stationary, it would lead to misleading results.

### 4.8 Filters

The estimated spectrum of a time series shows how variance of the series is distributed as a function of frequency. Depending on the purpose of analysis, some frequencies may be of greater interest than
others, and it may be helpful to reduce the amplitude of waves at other frequencies by statistically filtering them out before viewing and analyzing the series.

Filters transform time series into new time series so as to modify the data to meet particular objectives or display specific features of the data. The objective is to eliminate the power (of the spectrum) below a given frequency and to allow or pass the power above that value. Filters can be classified into three categories:

1. **Low pass filter** is a filter which passes low frequencies and cuts off high frequencies. The idea is to completely eliminate the power of the spectrum above a given frequency and allow the power below that frequency.

   *Example:* Smoothing is a form of filtering which produces a time series in which the importance of the spectral components at high frequencies is reduced.

2. **High pass filter** is a filter which passes high frequencies and cuts off low frequencies.

   *Example:* Detrending is a form of high-pass filtering. The fitted trend line tracks the lowest frequencies, with the result that the detrended series (i.e. residuals) would not have low frequencies.

3. **Band pass filter** is a filter that eliminates frequencies outside a certain range and allows frequencies within the given range.

### 4.9 Cross-spectrum

Cross-spectrum is an extension of cross-correlation to the frequency domain. It may be defined as the Fourier transform of cross-correlation between two time series.

Cross-spectral analysis is a powerful tool in the exploration of unknown relationships between two time series $x(t), y(t)$ where the correlation structure may vary over the time horizon. In effect, cross-spectral analysis decomposes two series in their individual cyclical components.

Cross-spectrum analysis is essentially concerned with the investigation of the joint structure of two time series, i.e. the dependence of one series on the other. However, one can only observe the relationships at the same frequency in both series. Let $S_x(\omega)$ and $S_y(\omega)$ denote the discrete Fourier transforms of the two series, where

$$S_x(\omega) = \frac{1}{N} \sum_{t=0}^{N-1} x(t) e^{-it\omega} \quad (40)$$

$$S_y(\omega) = \frac{1}{N} \sum_{t=0}^{N-1} y(t) e^{-it\omega} \quad (41)$$

The periodograms of these series are

$$S_{xx}(\omega) = |S_x(\omega)|^2 \quad \text{and} \quad S_{yy}(\omega) = |S_y(\omega)|^2 \quad (42)$$

The cross-periodogram is given by

$$S_{xy}(\omega) = \frac{N}{2\pi} \left[ S_x(\omega) S_y^*(\omega) \right] \quad (43)$$

The asterisk $*$ refers to the complex conjugate of the function, which is simply

$$\text{Conj}(y) = \text{Real}(y) - i\text{Imag}(y) \quad \text{for a complex function } y.$$
Notice that unlike the power spectrum, the cross-spectrum is complex valued as it contains amplitude and phase information. The real part of the cross-spectrum is known as the coincident spectrum (or co-spectrum), and the complex part of the cross-spectrum is known as the quadrature spectrum (or quad-spectrum).

One can estimate the phase and amplitude relationship between the two series from the cross-spectrum. Coherency between the two series is given by the following equation:

\[
S_{xy}(\omega) = \frac{|S_{xy}(\omega)|}{\sqrt{S_{xx}(\omega)S_{yy}(\omega)}}
\]  

(44)

The amplitude relationship is given by \(|S_{xy}(\omega)|^2\).

Coherency is a measure of correlation between two time series for each frequency \(\omega\). Coherency squared value of unity indicates complete dependence of one series on the other, whereas a coherency squared value of zero refers to no dependence of one series on the other.

The phase is calculated directly from the co-spectrum and quadrature spectrum as follows:

\[
\varphi_{xy}(\omega) = \tan^{-1} \left( \frac{\text{Imag}S_{xy}(\omega)}{\text{Real}S_{xy}(\omega)} \right)
\]  

(45)

Bivariate spectral analysis thus decomposes two series in individual cyclical components. For each period, the coherence is a measure of the highest possible squared correlation between the frequencies, \(\omega\), of the two series. The corresponding phase indicates at what lag of one of the series this correlation occurs. A high variability in the phase around a frequency may indicate that the corresponding coherence may be estimated with a lot of noise at that particular frequency.

Smoothing of spectral estimates is mandatory for the estimation of coherency. If no smoothing is performed, the estimate of coherency is identically equal to 1. Moreover, the estimation of phase is meaningless unless smoothing is performed.

**Interpretation of cross-spectrum**

Cross-spectral analysis allows examination of the similarities and co-movement of two separate time series in the frequency space. The coherency spectrum between two time series measures the predictability of one series from another at a given frequency. The phase spectrum provides evidence on the lead-lag relationship between two time series in the frequency space, thus evidence of causality may be derived from analyzing the phase spectrum in conjunction with the coherency spectrum. It indicates whether one time series is following the other and at what time lag.

**Summing up**

Spectral analysis involves the following sequence of steps:

- Detrending and pre-whitening,
- Windowing,
- Estimation of spectrum through FT or FFT.
Example 1: Visual inspection of time series plot

Dataset: AIRLINE

![Plot of Airline passengers’ data](image)

**Fig. 1.1: Plot of Airline passengers’ data**

**Comments**

The plot of monthly counts of Airline passengers’ data shows that:

- The number of passengers tends to increase over time (positive trend).
- The number of passengers tends to peak in certain months in each year.
- The spread or variance in the counts of passengers tends to increase over time.

**Interpretation**

- The series is not stationary, since the mean and variance tend to increase over time.
- The series also indicates a seasonal component.
Example 2: Statistics and Randomness tests

Dataset: NEWBOLD

![Time Series Plot](image1)

**Fig. 2.1:** Plot of time series NEWBOLD.DAT

<table>
<thead>
<tr>
<th></th>
<th>treasury</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean</td>
</tr>
<tr>
<td>2</td>
<td>Stand. dev.</td>
</tr>
<tr>
<td>3</td>
<td>Min</td>
</tr>
<tr>
<td>4</td>
<td>Max</td>
</tr>
</tbody>
</table>

**Fig. 2.2:** Descriptive statistics and Randomness test

![Descriptive Statistics Table](image2)

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Turning points</td>
<td>-1.596139</td>
<td>0.1104577</td>
</tr>
<tr>
<td>2. Daniel test</td>
<td>1.539189</td>
<td>0.1013023</td>
</tr>
</tbody>
</table>

![Autocorrelation Plot](image3)

**Fig. 2.3:** Autocorrelation plot

Comments

We can make the following conclusions from the above plots:

**Figure 2.1:** The time series plot does not show any increasing or decreasing trend. The values fluctuate around the mean. This is confirmed by the randomness test statistics and associated P-values.
Figure 2.3: There are no significant autocorrelations. Except for autocorrelation at lag 0, which is always equal to 1 by definition; all the autocorrelations fall within 95% confidence limits. In addition, there is no apparent pattern (such as a sequence of positive autocorrelations followed by a sequence of negative autocorrelations). This means that there is no associative ability to infer from a current value of the time series to the next value. Such non-association is the essence of randomness. In short, adjacent observations do not “co-relate”.

Randomness test statistics as well as the autocorrelation plot indicate that the series are random.

Example 3: Transformation - Moving Average smoothing

Dataset: AIRLINE

The plot of monthly counts of Airline passengers’ data shows periodic oscillations plus an increasing trend (see Example 2). It looks that the passenger counts tend to increase during summer months. Hence, we have smoothed the data using moving average window of 12 months. The results are plotted in Figure 3.1 below.

![Fig. 3.1: Plot of Airline passengers’ data after smoothing](image)

Comments

- The plot shows an increasing trend without any periodic peaks.
- The effect of smoothing can also be visualized by comparing the autocorrelation plots of smoothed and unsmoothed data. Figures 3.2 and 3.3 show Autocorrelation plots of Airline passengers’ data before and after smoothing.
Comments

- Notice the difference in the autocorrelation plots in Figures 3.2 and 3.3. The autocorrelation plot in Figure 3.3 is quite smooth compared to that in Figure 3.2.

- The plot starts with a high autocorrelation at lag 1 (only slightly less than 1) and uniformly declines to zero and then starts showing a negative trend. Such a pattern of autocorrelation is a signature of strong autocorrelation, which in turn provides high predictability of the smoothed time series.

- Without smoothing, the predictability of the time series would be poor.
Example 4: Cross-correlation

Dataset: SPNDMONY

The file SPENDMONY.DAT contains two time series measured simultaneous every three months

- Consumer Expenditure,
- Money Stock.

Fig. 4.1: Plot of time series Consumer Expenditure

Fig. 4.2: Plot of time series Money Stock

Comments

Figures 4.1 and 4.2 show that both series have increasing trend. In order to study cross-correlation between the two series, it would be essential to detrend the series. We have used the DIFFERENCE procedure (lag 1) to remove the trend.
Fig. 4.3: Autocorrelation and Cross-correlation plots of time series (raw data)

Comments

- Large positive cross-correlation at small time lags.
- Small positive cross-correlation at large time lags.
- There is no difference in the pattern of cross-correlation irrespective of whether Consumer Expenditure is selected first or Money Stock is selected first in the calculation of cross-correlation.

However, these results are misleading. Figure 4.4 shows a quite different pattern of cross-correlation between the detrended series.

Fig. 4.4: Autocorrelation and Cross-correlation plots of detrended time series
Comments

- When *Money Stock* is selected first in the calculation of cross-correlation, all cross-correlations are within 95% confidence band.

- When *Consumer Expenditure* is selected first in the calculation of cross-correlation, only one cross-correlation at time lag of 3 months is statistically significant. All other cross-correlations are within 95% confidence band.

- When *Consumer Expenditure* is selected first in the calculation of cross-correlation, *Consumer Expenditure* is negatively correlated with *Money Stock* 3 months later, but the converse is not true.

Example 5: Parametric Trend models

Dataset: MYFILE

![Graph showing an increasing trend](image)

**Fig. 5.1: Plot of the variable Investment**

Search for appropriate model

Figure 5.1 shows an increasing trend. All the parametric models (see section 3.2 above) were fitted to the data. The quadratic model issued the lowest residual variance and hence was selected for further analysis or elaboration.

\[ x_t = At^2 + Bt + C \]

The coefficients of the model issued by the program:

A = 0.007390823
B = 0.2123387
C = 72.03004

Parametric model is

\[ x_t = 0.007390823t^2 + 0.2123387t + 72.03004 \]
Fig. 5.2: Plot of the equation $x_t = 0.007390823t^2 + 0.2123387t + 72.03004$

Fig. 5.3: Plot of the detrended series
Example 6: Autoregression model

Dataset: PARTICLE

Fitting an Autoregression model involves the following sequence of steps:

1. Plotting of time series to visually check for any non-stationarity, outliers and seasonal patterns.
2. Transformation of data to induce stationarity.
3. Fitting of the model.

Step 1: Visual inspection of time series plot
Figure 6.1 indicates:

- Strong and positive correlation,
- Absence of any significant seasonal variations.

Figure 6.2 shows:

- Strong and positive autocorrelations,
- Autocorrelation decays very slowly,
- The process generating the time series is not stationary.

**Step 2: Transformation of the data**

Stationarity was induced by differencing with lag 1. Figure 6.3 shows the plot of the differenced data.
It can be easily seen from Figure 6.3 that:

- The mean of the differenced data is around zero.
- Autocorrelation is less strong than in the original data.

In order to confirm the last observation, Autocorrelation was performed on the differenced data.

The autocorrelation plot shows that only one autocorrelation is significant at 5% level. Figures 6.3 and 6.4 together indicate that the differenced data are stationary, and hence appropriate for fitting the Autoregression model.

**Step 3: Model fitting**

A series of Autoregression models were fitted with order ranging from 1 to 4. Figure 6.5 shows a plot of error variance for different values of model order. A model of order 2 was selected for further analysis and elaboration.
Example 7: Spectral analysis of random time series

Dataset: RANDOM

Pre-processing

Since the series is random, there is no need for pre-processing.
Comments

We can make the following conclusions from the spectral plot:

- There are no dominant peaks.
- There is no identifiable pattern and the peaks seem to fluctuate at random.

The spectral plot is a signature of random data.

Example 8: Spectral analysis (Maximum entropy and Fourier analysis options)

Dataset: RALEIGH

Pre-processing

- *Demeaning*: The mean of the series was subtracted by invoking the paired arithmetic option in the transformation module.

- *Checking for stationarity*: The autocorrelation plot in Figure 8.1 is a signature of periodicity in the time series. Randomness tests indicate non-randomness of the time series. The series was whitened by differencing twice prior to spectral analysis.
Spectral analysis - Maximum entropy option

Spectral plot and table of values (only a small relevant portion) issued by the program are given below.
<table>
<thead>
<tr>
<th>Frequency</th>
<th>Period</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.007692308</td>
<td>0.0006404413</td>
</tr>
<tr>
<td>2</td>
<td>0.01538462</td>
<td>0.0006526862</td>
</tr>
<tr>
<td>3</td>
<td>0.02307692</td>
<td>0.0006688896</td>
</tr>
<tr>
<td>4</td>
<td>0.03076923</td>
<td>0.0005793989</td>
</tr>
<tr>
<td>5</td>
<td>0.03846154</td>
<td>0.0005207116</td>
</tr>
<tr>
<td>6</td>
<td>0.04615385</td>
<td>0.0005156625</td>
</tr>
<tr>
<td>7</td>
<td>0.05384615</td>
<td>0.0006599957</td>
</tr>
<tr>
<td>8</td>
<td>0.06153846</td>
<td>0.0009051206</td>
</tr>
<tr>
<td>9</td>
<td>0.06923077</td>
<td>0.001928888</td>
</tr>
<tr>
<td>10</td>
<td>0.07692308</td>
<td>1.00000000</td>
</tr>
<tr>
<td>11</td>
<td>0.08461539</td>
<td>0.000848084</td>
</tr>
<tr>
<td>12</td>
<td>0.09230769</td>
<td>0.000724181</td>
</tr>
<tr>
<td>13</td>
<td>0.1076923</td>
<td>0.0009565167</td>
</tr>
<tr>
<td>14</td>
<td>0.1153846</td>
<td>0.0005582218</td>
</tr>
<tr>
<td>15</td>
<td>0.1230769</td>
<td>0.0003981288</td>
</tr>
<tr>
<td>16</td>
<td>0.1307692</td>
<td>0.0003515273</td>
</tr>
</tbody>
</table>

Fig. 8.3: Table of spectrum values

Comments

We can make the following conclusions from the spectral plot and table of values:

- There is a single dominant peak at frequency approximately equal to 0.08 or time period of about 12 units (corresponding to the highest spectral density = 1.0).
- The underlying process is sinusoidal of single frequency.

Fourier analysis (Welch window)

Spectral plot and table of values (only a small relevant portion) issued by the program are given below.

Fig. 8.4: Spectral plot (Welch window)
<table>
<thead>
<tr>
<th>Frequency</th>
<th>Period</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2.35926400E-33</td>
</tr>
<tr>
<td>2</td>
<td>0.007692308</td>
<td>0.0002284034</td>
</tr>
<tr>
<td>3</td>
<td>0.01538462</td>
<td>0.001626399</td>
</tr>
<tr>
<td>4</td>
<td>0.02307692</td>
<td>0.001133167</td>
</tr>
<tr>
<td>5</td>
<td>0.03076923</td>
<td>0.0001322902</td>
</tr>
<tr>
<td>6</td>
<td>0.03846154</td>
<td>0.00005408447</td>
</tr>
<tr>
<td>7</td>
<td>0.04615385</td>
<td>0.0002777412</td>
</tr>
<tr>
<td>8</td>
<td>0.05384615</td>
<td>0.001424402</td>
</tr>
<tr>
<td>9</td>
<td>0.06153846</td>
<td>0.00312759</td>
</tr>
<tr>
<td>10</td>
<td>0.06923077</td>
<td>0.005025022</td>
</tr>
<tr>
<td>11</td>
<td>0.07692308</td>
<td>0.08706709</td>
</tr>
<tr>
<td>12</td>
<td>0.08461539</td>
<td>0.7744586</td>
</tr>
<tr>
<td>13</td>
<td>0.09230769</td>
<td>0.06299494</td>
</tr>
<tr>
<td>14</td>
<td>0.1</td>
<td>0.01055351</td>
</tr>
<tr>
<td>15</td>
<td>0.1076923</td>
<td>0.002736243</td>
</tr>
<tr>
<td>16</td>
<td>0.1153846</td>
<td>0.000806794</td>
</tr>
<tr>
<td>17</td>
<td>0.1230769</td>
<td>0.0004511417</td>
</tr>
<tr>
<td>18</td>
<td>0.1307692</td>
<td>0.000005709783</td>
</tr>
<tr>
<td>19</td>
<td>0.1384615</td>
<td>0.001426795</td>
</tr>
</tbody>
</table>

Fig. 8.5: Table of spectrum values

Comments

- We get the same substantive results, but the peak is not as sharp as in the case of Maximum entropy option.

- There is a single dominant peak at frequency approximately equal to 0.08 (corresponding to the highest spectral density (=0.7744586)).

Fourier Analysis (No window)

![Spectral plot (No window)](image)

Fig. 8.6: Spectral plot (No window)
<table>
<thead>
<tr>
<th>Frequency</th>
<th>Period</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2.35926400E-33</td>
</tr>
<tr>
<td>2</td>
<td>0.007692308</td>
<td>0.0002284034</td>
</tr>
<tr>
<td>3</td>
<td>0.01538462</td>
<td>0.001626399</td>
</tr>
<tr>
<td>4</td>
<td>0.02307692</td>
<td>0.001133167</td>
</tr>
<tr>
<td>5</td>
<td>0.03076923</td>
<td>0.0001322902</td>
</tr>
<tr>
<td>6</td>
<td>0.03846154</td>
<td>0.00005408447</td>
</tr>
<tr>
<td>7</td>
<td>0.04615385</td>
<td>0.0002777412</td>
</tr>
<tr>
<td>8</td>
<td>0.05384615</td>
<td>0.001424402</td>
</tr>
<tr>
<td>9</td>
<td>0.06153846</td>
<td>0.00312759</td>
</tr>
<tr>
<td>10</td>
<td>0.06923077</td>
<td>0.005025022</td>
</tr>
<tr>
<td>11</td>
<td>0.07692308</td>
<td>0.08706709</td>
</tr>
<tr>
<td>12</td>
<td>0.08461539</td>
<td>0.7744586</td>
</tr>
<tr>
<td>13</td>
<td>0.09230769</td>
<td>0.06299494</td>
</tr>
<tr>
<td>14</td>
<td>0.1</td>
<td>0.01055351</td>
</tr>
</tbody>
</table>

Fig. 8.7: Table of spectrum values

Comments
The results are exactly the same as in the case of Welch window.

Example 9: Cross-spectrum

Dataset: ASTHMA

Two Variables: Number of deaths due to bronchial asthma (1) Male (2) Female.

Pre-processing

_Demeaning:_ Both the series were demeaned by subtracting their respective means invoking the Paired Arithmetic option in the Transform menu.

_Checking for stationarity:_ The autocorrelation plots in Figures 9.1 and 9.2 are signatures of periodicity in the time series for both Male and Female.

_Randomness tests_ indicated the series were not random. The series were whitened by differencing twice prior to spectral analysis.

![Fig. 9.1: Autocorrelation plot of demeaned series (Male)](image)
Fig. 9.2: Autocorrelation plot of demeaned series (Female)

Cross-spectrum

Fig. 9.3: Coherency spectrum

Fig. 9.4: Phase spectrum
Comments

Coherency: It can be easily seen from Figure 9.3 that coherency is maximum at low frequencies but remains substantial throughout the frequency range. The coherency ranges between 0.371 and 0.974.

Phase: The phase spectrum plotted in Figure 9.4 does not show any definite trend (neither increasing nor decreasing). It rather fluctuates around zero phase. The data are compatible with an underlying zero phase spectrum.

Interpretation

The combination of zero phase spectrum and strong coherency suggest that the fluctuations in the two whitened series are largely the result of fluctuations in some common external factor (perhaps Environment).