Section 2 - Time Series Analysis: Weeks 8-13

Lecturer: Associate Professor Shelton Peiris, Carslaw 819 - Ext 15764

General Information for Time Series Analysis - Section 2

In 2013 the lectures on Time Series Analysis section (Section 2 of STAT3013/3913) will begin in Week 8. Therefore, the first lecture will be given on Monday April 29. The weekly tutorials and practicals will begin on the Friday May 3 in Week8. Classes end on Friday 7 June. All classes are in common for both STAT3011 and STAT 3911. No additional material or lectures will be given for STAT3911 Section 2 (Time Series) of the course. All electronic material will posted on the STAT3011 homepage only.

Objectives: Establish some methods of modelling and analysing (ie. identification, estimation, decision making, and prediction) of time series data (ie. data containing some serially dependence structure). Some real world applications will be discussed. The package R will be used to analyse time series data.

Outcomes: Successful completion of this unit, students will be able to:

(1) Identify a time series and its various components,
(2) Apply various transformations to smooth a time series,
(3) Identify Stationary and homogeneous non-stationary time series,
(4) Autocorrelation and partial autocorrelation functions,
(5) Identify suitable ARMA and ARIMA models for given time series data,
(6) Parameter estimation,
(7) Diagnostic checking procedures and forecasting,
(8) Apply the statistical package R to model and forecast time series data.
# Course Outline for the Time Series Component

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Time series data, components of a time series. Filtering to remove trends and seasonal components.</td>
</tr>
<tr>
<td>9</td>
<td>Stationarity time series. Sample autocorrelations and partial autocorrelations. Probability models for stationary time series. Moving Average (MA) models and properties.</td>
</tr>
<tr>
<td>10</td>
<td>Invertibility of MA models. Autoregressive (AR) models and their properties. Stationarity of AR models. Mixed Autoregressive Moving Average (ARMA) models and their properties.</td>
</tr>
<tr>
<td>12</td>
<td>Estimation and fitting ARIMA models via MM and MLE methods. Hypothesis testing, diagnostic checking and goodness-of-fit tests. AIC for ARIMA models. Forecasting methods for ARIMA models.</td>
</tr>
</tbody>
</table>

## Assessment

Section 2 - Time Series Analysis - 50% as given below:

- 1 Assignment (due Friday 17 May) 3%
- 1 tutorial quiz (Friday 31 May) 7%
- 5 Computer practicals (weeks 8/9-12/13) 5%
- Final examination in June 35%

## References


***************************************************************************************
1 Basic Concepts in Time Series

A time series is a set of observations taken sequentially at specified times.

Examples of time series occur in variety of fields ranging from economics to engineering.

Examples:


The above data are given below:

\[(1)\]
464 675 703 887 1139 1077 1318 1260 1120 963 996 960 530 883 894 1045 1199 1287 1565 1577 1076 918 1008 1063 544 635 804 980 1018 1064 1404 1286 1104 999 996 1015 615 722 832 977 1270 1437 1520 1708 1151 934 1159 1209 699 830 996 1018 1064 1404 1286 1104 999 996 1015 615 722 832 977 1270 1437 1520 1708 1151 934 1159 1209 699 830 996 1018 1064 1404 1286 1104 999 996 1015 615 722

\[(2)\]

\[(3)\]
To see more time series data, visit:


Note:

1. The observations in time series may take hourly, daily, weekly, monthly, quarterly, annually, etc. according to the problem of interest.

2. If the data observed at equally spaced time (or regular time intervals), then the time series is called an equally spaced discrete time series or a regular time series (rts). However, there are many situations arise in practice that the time scale is considered as continuous. This type of continuous time series occur when the interval between successive observations is very small or $\Delta t$. However, this course concentrates the analysis of discrete, regular time series and investigate their important statistical properties.

1.1 A notation for a time series data

$N$ readings (or observations) of a time series are denoted by $x_1, x_2, \cdots, x_N$, and written as \{x_t\}, $t = 1, \cdots, N$. Each $x_t$ can be considered as a dependent function of time and hence a suitable deterministic (or non-stochastic) model for $x_t$ is given by $x_t = f(t)$, where $f(t)$ is a suitably chosen deterministic function of time depending on its behaviour. For example, $f(t)$ may be a polynomial in $t$ such that

- $f(t) = a + bt$ (for linear behaviour)

- $f(t) = at^2 + bt + c$ (for quadratic behaviour) or any other function of $t$ satisfying

- $f(t) = a \sin t + b \cos t$ (for oscillating behaviour) etc.
However, from statistical point of view observed numerical data $x_1, \cdots, x_N$ can be thought of as observations on sequence of random variables $X_1, \cdots, X_N$ which are, in general, not independent. Thus a nondeterministic or a simple stochastic model for $X_t$ is given by $X_t = f(t) + e_t; \quad t = 1, 2, \cdots, N$ where $e_t$ is an unobservable random noise (or error) term at time $t$. For simplicity, it is assumed that $e_t$ has zero mean and constant variance. Using the standard notation we write this as $E(e_t) = 0$ and $\text{Var}(e_t) = \text{constant}(\sigma^2)$ for all $t$. In general, $\{e_t\}$ is considered as a sequence of uncorrelated random variables which are not necessarily independent. Recall that independence implies the variables are uncorrelated, however, the opposite is not true unless the variables are from a normal distribution.

In practice, we observe time series superimposed with noise as follows:

The stochastic analysis of time series is concerned with the analysis of dependence structure and fitting a suitable model to describe its long term behaviour. This has two steps: identification and parameter estimation. The next step of the analysis is the diagnostic checking using a suitable test to see the adequacy of the fitted model. If the fitted model is satisfactory then forecasts of future values can be obtained with less uncertainty. Thus the analysis of time series consist of 4 steps, namely, (i) identification, (ii) parameter estimation, (iii) diagnostic checking and, and (iv) forecasting.
1.2 Graphs of Time Series Data: Time Series Plots

The first step in time series analysis is to plot the time series data against their corresponding time points to observe the behaviour of the series. This is called a time series plot (ts plot).

**Example:** Plot the time series given by

<table>
<thead>
<tr>
<th>Time, $t$</th>
<th>1 2 3 4 5 6 7 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value, $x_t$</td>
<td>8 4 4 3 1 1 1 1</td>
</tr>
</tbody>
</table>

**Solution:** Plot the above points $x_t$ against $t$.

This time series plot (tsplot) gives us a clear picture of the short and/or long term behaviour of the series. Time series plots are used at the preliminary stage of analysis. Many modern computer packages can be used in time series analysis. In this course we use the package R to analyze some time series data.

The \textit{R} command to produce a \textit{ts} plot for a given set of time series data is:

```r
> ts.plot (data)
```

Note: This command assumes the data as a regular time series.

**Exercise**

1. Investigate the tsplots 1 to 3 and comment.

2. Draw a tsplot in the space below for the time series:

<table>
<thead>
<tr>
<th>Time, $t$</th>
<th>1 2 3 4 5 6 7 8 9 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value, $x_t$</td>
<td>40 45 35 65 70 55 50 75 80 45</td>
</tr>
</tbody>
</table>

These \textit{ts} plots give clear pictures of each time series about their long term behaviour.

(Also see pp. 1-3 of Chatfield for more information.)
R Commands

> par(mfrow=c(3,1))

> ts.plot (data1)

> ts.plot (data2)

> ts.plot (data3)

Exercise: Comment on each time series (see Sheet 1).

1.3 Some Typical Time Series Plots

1) Constant Mean: The mean level of a ts does not change with time. In practical situations the series fluctuates slowly around a constant level and hence the change in the mean level is negligible.

2) Linear Trend ($T_t$): This is the general direction in which the graph of a ts appears to be going over a long interval of time.

   **Upward linear trend**

   ![Upward linear trend graph]

   **Downward linear trend**

   ![Downward linear trend graph]
3) Cyclical Movements ($C_t$) : Cyclical variations refer to the long-term oscillations or swings about a trend line. They may or may not follow exactly similar patterns after equal intervals of time (i.e. the series may or may not be periodic).

4) Seasonal Movements ($S_t$)

Seasonal movements or seasonal variations refer to identical or almost identical patterns of a time series appears to follow in succession in a long run.

Note: This time interval for identical patterns is called the periodicity. Seasonal movements in general refer to annual periodicity. However, this can be extended to include periodicity over any interval of time such as hourly, daily, weekly, etc.

Remark : It is clear that all seasonal variations are cyclical but not vice-versa.
5) Irregular or Random Movements ($I_t$)

This refers to the sporadic motions of time series due to chance events such as floods, strikes, elections etc. We usually assume that such events produce variations lasting only a short time. These events may result in new cyclical or other movements. (Also see Ch.2 of Chatfield.)

![Irregular variations](image)

### 1.4 Basic Terminology of Time Series Analysis

An investigation of the components $T, C, S$, and $I$ of a time series is often referred to as the decomposition of a $ts$ into its basic components. In general,

- a multiplicative model with these 4 components is given by

$$X_t = T_t \cdot C_t \cdot S_t \cdot I_t$$

  to represent the data point at time $t$.

- an additive model is given by

$$X_t = T_t + C_t + S_t + I_t.$$

Notice that any multiplicative model can be linearised by taking a logarithmic transformation.

**Some Transformations of Time Series Data**

A $ts$ plot of the data may indicate that it is sensible to consider suitable transformation for further analysis. There are several standard transformations available in practice. If the variance appears to increase with the mean, then try a logarithmic transformation to stabilise the variance.

2 diagrams - Before & After the ‘log’ transformation.

**Note:** If there is a trend and the size of the seasonal effect appears to increase with the mean then use a logarithmic transformation to make the seasonal effect additive.
The corresponding R Command is:

\[ \log(\text{data}) \text{ (for natural logarithm)} \quad \text{or} \quad \log_{10}(\text{data}) \quad \text{(for standard, base 10 logarithmic) transformations.} \]

**Stationary Time Series**

In many practical problems in time series analysis, one can transform the series such that there is no systematic change in mean level (for example, no trend) and no systematic change in the variance. Such stable (in both mean and variance) time series often called a stationary time series.

Loosely speaking, a time series \( \{X_t\} \) is said to be stationary if it has similar statistical properties for all values of \( t \). That is, \( E[X_t] \) and \( \text{Var}[X_t] \) are constant or independent of \( t \).

**Note:** This is not a proper definition of a stationary ts and the series need another strong condition to be satisfied. However, the ‘stationarity’ of a ts can be recognised in this way by investigating the mean and variance. We’ll discuss stationary ts in detail later in this course and, further in an advanced level course.

### 1.5 Analysis of Components of a Time Series

As we have mentioned before, the first step in the analysis of ts is to plot the data. A careful inspection of this graph may suggest the use of a multiplicative or additive model to data. For example, an additive model is \( X_t = T_t + S_t + e_t \); where \( e_t \) is the random error term at time \( t \) with \( E(e_t) = 0 \) and \( \text{Var}(e_t) = \sigma^2 = \text{const.} \) This error term \( e_t \) is also called a noise component. An alternative multiplicative model is \( X_t = T_t \cdot S_t \cdot e_t \)

**Note:** \( T_t \) and \( S_t \) are non-random functions of \( t \).

i.e. \( E(T_t) = T_t \); \( E(S_t) = S_t \) and \( \text{Var}(T_t) = \text{Var}(S_t) = 0. \)

Now clearly, for an additive model, \( E(X_t) = T_t + S_t \) and hence \( \mu_t = T_t + S_t \), where \( E(X_t) = \mu_t. \) However, \( \text{var}(X_t) = \sigma^2. \) The components \( T_t \) and \( S_t \) are non-random and \( S_t \) is assumed to have period \( d \) (i.e. identical or almost identical patterns occur after every \( d \) time points). In other words \( d \) is the smallest positive integer such that \( S_{t+d} = S_t \).

**Note:** Clearly, the t.s. given by \( X_t = T_t + S_t + e_t \) is not stationary in the sense that its mean is not constant. Thus we can transform \( \{X_t\} \) into a stationary ts by removing \( T_t + S_t \) from \( X_t \) for each \( t \), i.e. \( X_t - T_t - S_t \) gives a stationary series.

**Examples:**

1) \( S_t = \sin t \)

\[ S_{t+2\pi} = S_{t+4\pi} = \cdots = \sin t \]

Thus the period, \( d = 2\pi. \) In general, \( S_t = \sin kt \) has period, \( d = 2\pi/k. \)

2) Annual periodicity with

(a) monthly observations, \( d = 12, \)

(b) 4-weekly observations, \( d = 13, \) etc.
Note: Including the cyclical component, \( C_t \), we may write a general additive model for the data as \( X_t = T_t + S_t + C_t + e_t \) where \( \mu_t = E(X_t) = T_t + S_t + C_t \) is changing with time (i.e. not a constant). Therefore it is also not a stationary time series.

Now we examine some techniques for estimating and eliminating the components \( T_t \) and \( S_t \) from a given \( ts \) with no cyclical variations. **Estimation and Elimination of Trend and Seasonal Components**

### A. Estimation and Elimination of Trend in the Absence of Seasonality.

Model: \( X_t = T_t + e_t; \ t = 1, 2, \ldots, n \)

**Case 1. Fitting a suitable polynomial.**

i) Linear Trend:

\[ T_t = \alpha + \beta t \]

estimate \( \hat{T}_t \) using the method of l.s., then remove \( \hat{T}_t \) from \( x_t \).

Using the method of l.s, it is easy to see that

\[ \hat{\beta} = \frac{\sum(x_i - \bar{x})(t_i - \bar{t})}{\sum(x_i - \bar{x})^2} \]

and \( \hat{\alpha} = \bar{x} - \hat{\beta}\bar{t} \).

Now, the estimated residuals \( \hat{e}_t = x_t - \hat{T}_t \) gives a series with no linear trend.

ii) Polynomial Trend:

\[ T_t = \alpha_0 + \alpha_1 t + \cdots + \alpha_p t^p \]

estimate \( \hat{T}_t \) using the method of l.s.

Now remove the trend \( \hat{T}_t \) from \( x_t \). i.e. \( x_t - \hat{T}_t \) gives a series with no polynomial trend.

**Note:** Since the series \( \{x_t - \hat{T}_t\}; \ t = 1, 2, \ldots, n \) has no trend, it is called a detrended \( ts \).

**Exercise:** Find the estimates of \( \alpha_0, \alpha_1; \) and \( \alpha_2 \) of (ii) when \( p = 2 \) using the method of l.s.

**Case 2. Using a Suitable Moving Average (MA) Filter**

Suppose that we have a \( ts \) with a trend \( T_t \). Assume that \( T_t \) is approximately linear over the interval \([t - q, t + q]\), where \( q \) is a suitably chosen integer (after careful inspection of the \( ts \) plot). Let \( d = 2q + 1 \) (an odd number) be the length of the interval.

E.g. If \( q = 2 \), then \( d = 5 \)

Consider the data set \( x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10} \).

We average the every 5-consecutive values of \( \{x_t\} \) and obtain

\[
\begin{align*}
   y_1 &= \hat{T}_1 = \text{cannot calculate} \\
   y_2 &= \hat{T}_2 = \text{cannot calculate} \\
   y_3 &= \hat{T}_3 = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} \\
   y_4 &= \hat{T}_4 = \frac{x_2 + x_3 + x_4 + x_5 + x_6}{5} \\
   &\vdots \\
   y_8 &= \hat{T}_8 = \frac{x_6 + x_7 + x_8 + x_9 + x_{10}}{5}
\end{align*}
\]
Clearly there are only 6 values in this list, and "lost" two values at each end since we have no sufficient information to calculate them. This procedure can be written as

\[
y_t = \hat{T}_t = \frac{\sum_{r=-2}^{2} x_{t+r}}{5}; \quad t = 3, 4, \cdots, 8
\]

In general, the linear trend corresponding to the middle time point of the span is estimated by

\[
y_t = \hat{T}_t = \frac{1}{2q + 1} \sum_{r=-q}^{q} x_{t+r}; \quad t = q + 1, \cdots, N - q
\]

notice that we “lost” 2q observations (i.e. q observations at each end) in this process.

This procedure is called a \textit{ma} (or simple \textit{ma}) of length or span 2q + 1.

Now the detrended series is obtained by \( x_t - \hat{T}_t; \quad t = q + 1, \cdots, N - q \)

**Numerical Example 1:** Construct an \textit{ma} of span 5 from the data below:

<table>
<thead>
<tr>
<th>( x_t )</th>
<th>2</th>
<th>1</th>
<th>5</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>6</th>
<th>10</th>
<th>3</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>total:</td>
<td>*</td>
<td>*</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>30</td>
<td>35</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>( ma: )</td>
<td>*</td>
<td>*</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

**Numerical Example 2:** Draw the ts plot and its \textit{ma} of above data

Notes:

(1) Using the above procedure one can smooth out the local fluctuations and estimate the local mean of the series. Thus \( y_t = \hat{m}_t \) is called a smoothed value of \( x_t \) (or \( sm(x_t) \)) [See p.14 of \textit{Chatfield}]

\[
sm(x_t) = y_t = \hat{T}_t = \frac{1}{2q + 1} \sum_{r=-q}^{q} x_{t+r} \quad \text{is called filtering.}
\]

The weights of this \textit{ma} filter are \( \frac{1}{2q + 1} \) (equal weights).

(2) When we use a filter of the above type, there is likely to be an end-effect problem, since \( sm(x_t) \) is calculated for \( t = q + 1 \) to \( t = N - q \).
The R command:

```r
> filter(data, rep(1/(2q+1), 2q + 1))
```
e.g. when $q = 2$ (for a ma of span 5)

```r
> filter(x, rep(1/7, 5))
```
gives the ma values in row 3.

**Centered Moving Averages**

When the length of a required ma is an even number, the ma’s (or the averages), do not correspond to any time point as in the odd case before.

e.g. when $d = 4$

e.g.

\[
\begin{align*}
\text{y}_1 &= \frac{x_1 + x_2 + x_3 + x_4}{4} \\
\text{y}_2 &= \frac{x_2 + x_3 + x_4 + x_5}{4}
\end{align*}
\]

correspond to the averages between $(t_2, t_3), (t_3, t_4), \ldots$ or at $t_{2.5}, t_{3.5}, \ldots$

A sensible way to avoid the above mis-representation is to average the values $y_1$ and $y_2$ again to represent the value at $t_3$. Clearly, the average of $y_1$ and $y_2$ or $\frac{y_1 + y_2}{2}$ gives a good estimate at $t_3$. This procedure is known as the **Centered Moving Averages**.

**An Illustration**

Suppose that the length (or span) of a ma is 4 or $d = 4$

In this case we use the following procedure.

\[
\begin{align*}
\text{4-year} & \quad \text{4-year Centered ma} \\
\text{monthly total} & \quad \rightarrow (\text{NA}) \\
x_1 & \rightarrow (\text{NA}) \\
\rightarrow x_1 + x_2 + x_3 + x_4 & \rightarrow \frac{1}{8}(x_1 + 2x_2 + 2x_3 + 2x_4 + x_5) \\
x_3 & \rightarrow x_2 + x_3 + x_4 + x_5 \\
\rightarrow x_2 + x_3 + x_4 + x_5 & \rightarrow \frac{1}{8}(x_2 + 2x_3 + 2x_4 + 2x_5 + x_6) \\
x_4 & \rightarrow x_3 + x_4 + x_5 + x_6 \\
\rightarrow x_3 + x_4 + x_5 + x_6 & \rightarrow \frac{1}{8}(x_3 + 2x_4 + 2x_5 + 2x_6 + x_7) \\
x_5 & \rightarrow \frac{1}{8}(x_3 + 2x_4 + 2x_5 + 2x_6 + x_7) \\
x_6 & \\
x_7 & \\
\vdots
\end{align*}
\]

The last column is called a 4-year centered ma, since it centers two successive averages at $t = 3$. 

13
**Example:** Complete the table below using a centered \( ma \) filter of length (or span) 4. Plot the smoothed series.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( x_t )</th>
<th>Moving sum of 4</th>
<th>Centered ma of span 4 = ( sm(x_t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>( 4 + 5 = 4.5 ) or ( 4.5 = \frac{1}{8}(1) + \frac{1}{4}(3 + 5 + 1) + \frac{1}{8}(5) )</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

**Note:** If the series has a linear trend, then after taking a suitable \( ma \) (ordinary or centered) remains the linear trend.

**The R command:**

In order to obtain a centered \( ma \) one need to specify the respective weights as given below:

a) for centered \( ma \) of span 4, the weights are \( \frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8} \) and hence the R command is:

\[
> \text{filter(data, c(1/8, 1/4, 1/4, 1/4, 1/8))}
\]

b) for a centered \( ma \) of span \( q \) (\( q \)-even) the weights are \( \frac{1}{2q}, \frac{1}{q}, \frac{1}{q}, \cdots, \frac{1}{q}, \frac{1}{2q} \), \( q-1 \) times.

\[
> \text{filter(data, c(1/(2q), 1/q, \cdots, 1/q, 1/(2q))) or} \\
> \text{filter (data, c(1/(2q), rep(1/q, q-1), 1/(2q)))}
\]

**Case 3. Elimination of Trends by Differencing.**

\( X_t - X_{t-d}; \, t = d + 1, \cdots, N \) is called lag \( d \) differencing.

i) Linear Trend

Model: \( X_t = \alpha + \beta t + \epsilon_t; \, t = 1, 2, \cdots, N \)
Consider the lag 1 differencing $X_t - X_{t-1}$

Clearly $X_t - X_{t-1} =$

Thus $Y_t = X_t - X_{t-1}$ has no trend and therefore $\{Y_t\}$ is stationary.

Notice: $Y_1 = NA$

$Y_2 = X_2 - X_1$, $Y_3 = X_3 - X_2, \ldots, Y_N = X_N - X_{N-1}$

**Example:**

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_t$</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>10</td>
<td>15</td>
<td>18</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>$y_t$</td>
<td>NA</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>-3</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Notice:

1) No. of observations in the original series = 10
   
   No. of observations in the final (or differenced) series = 9.
   
   Thus we “lost” one observation in differencing.

2) This procedure is called lag 1 or first-order differencing.

ii) Quadratic Trend

Model: $X_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t$; $t = 1, 2, \ldots, N$

Now the lag 1 differencing, $Y_t = X_t - X_{t-1} =$

Therefore, the series $Y_t$ has a linear trend. Then difference the data once again to remove this linear trend:

$U_t = Y_t - Y_{t-1} =$

Thus clearly $U_t$ has no trend and therefore $\{U_t\}$ is stationary.
Example:

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_t$</td>
<td>NA</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>-3</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$y_t$</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>-4</td>
<td>2</td>
<td>-6</td>
<td>8</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: There are only 8 observations in the differenced series. Thus we “lost” 2 observations in this second-order (or twice) differencing.

In R: $d1 = \text{diff (data)}$ (lag 1 diff.)

$d2 = \text{diff (d1)}$

**B. Estimation and Elimination of Both Trend and Seasonal Components of a TS**

Model: $X_t = T_t + S_t + \epsilon_t; t = 1, 2, \cdots, N$

Assume: $E(\epsilon_t) = 0, S_{t+d} = S_t$ and $\sum_{j=1}^d S_j = \text{constant}$, where $d$ is the period.

**Case I:** $d$ is an odd number ($d = 2q + 1$). In this case take a simple moving average of span $d$

i.e. $\frac{1}{2q+1} \sum_{r=-q}^{q} x_{t+r} = \frac{1}{2q+1} \sum_{r=-q}^{q} T_{t+r} + \frac{1}{2q+1} \sum_{r=-q}^{q} S_{t+r} + \frac{1}{2q+1} \sum_{r=-q}^{q} \epsilon_{t+r}$

$t = q + 1, \cdots, N - q$

Since $\sum_{r=-q}^{q} S_{t+r} = \text{constant}$, we have $y_t = \frac{1}{2q+1} \sum_{r=-q}^{q} x_{t+r} = sm(x_t); t = q + 1, \cdots, N - q$.

**Notes:**

(a) $q$ readings are “lost” at each end.

Clearly, $sm(x_t) = y_t = \hat{T}_t$ is an estimate of $T_t$ for $t = q + 1, \cdots, N - q$.

(b) Since $x_t - y_t = \hat{S}_t + \hat{e}_t; t = q + 1, \cdots, N - q$ we compute the seasonal component plus error using $\hat{s}_k + \hat{e}_k = x_k - y_k; k = 1, 2, \cdots, d$

**Case II:** If $d$ is an even number, then we use a suitable centered moving average of span $d$.

**Elimination of Seasonal Components by Differencing**

The technique of differencing to nonseasonal data can be adapted to deal with seasonality of period $d$ by introducing lag $d$ differencing,

i.e. Calculate $V_t = X_t - X_{t-d}; t = d + 1, \cdots, N$

Reason: $X_t = T_t + S_t + \epsilon_t$ and $s_t \approx s_{t-d}$

$$V_t = X_t - X_{t-d} = (T_t + S_t + \epsilon_t) - (T_{t-d} + S_{t-d} + \epsilon_{t-d})$$

$$= T_t - T_{t-d} + \epsilon_t - \epsilon_{t-d}, \text{ since } S_t \approx S_{t-d}$$

Since $S_t \approx S_{t+d} \approx s_{t-d}$, $V_t = X_t - X_{t-d}$ is free from the seasonal component. The trend component can be eliminated using the method’s already described before.
Seasonality with a Linear Trend

Let \( X_t = S_t + T_t + e_t \).

Now \( V_t = T_t - T_{t-d} + S_t - S_{t-d} + e_t - e_{t-d} \).

Exercise: Show that when \( T_t = \alpha + \beta t \), \( V_t = \beta d + e_t - e_{t-d} \).

Solution:

\[
V_t = (\alpha + \beta t) - [\alpha + \beta(t - d)] + e_t - e_{t-d} \\
= \beta d + e_t - e_{t-d} \text{ since } S_t \approx S_{t-d}
\]

i.e. If there is a linear trend in addition to the periodicity, \( X_t - X_{t-d} \) (lag d differencing) gives a deseasonalized, detrended series.

Elimination of Trend and Seasonal Components by Differencing (in practice)

Recall: For a nonseasonal time series \( x_t \) with a

a) linear trend use \( u_t = x_t - x_{t-1} \) (first diff. to remove it)

b) quadratic trend use \( u_t - u_{t-1} \) (2nd difference) to remove it.

In general \( x_t - \hat{T}_t - \hat{S}_t \) gives a stationary noise sequence. In many practical problems related to time series data, this noise is not a sequence of uncorrelated random variables. If the noise sequence does have sample autocorrelations significantly different from zero, then we can model this noise sequence as a member of the popular autoregressive integrated moving average (ARIMA) family. Accurately identified and fitted model for a given data set can be used to forecast future values in terms of the past values. This key role of autocorrelations in \( tsa \) is analysed by processes whose properties do not vary with time (or stationary time series).

Now we consider the stationary processes and their properties in detail.