Lecture 8

TIME SERIES AND FORECASTING

Introduction

A **time series** is a collection of observations made sequentially in time.

Time series can be **continuous** and **discrete**.

In this course we consider only discrete time series, with observations being taken at equally spaced intervals.

The first step in the analysis of time series is to plot the data against time in a **time series plot**.

The lecture notes give the commands for producing time series plots in `Minitab`. Note that the axes should be labelled appropriately and that the plot should have a title.

You should be able to **describe** a time series in words:

- Is there any **trend** in the plot? Is this trend “upwards” (increasing/positive) or “downwards” (decreasing/negative)?
- Is there any **seasonality** in the plot? Are there other, more complex, **cyclic patterns**?
- Does the plot show **both** trend **and** seasonality?
- Are there any **outliers**?
- Does the plot look **stationary**?

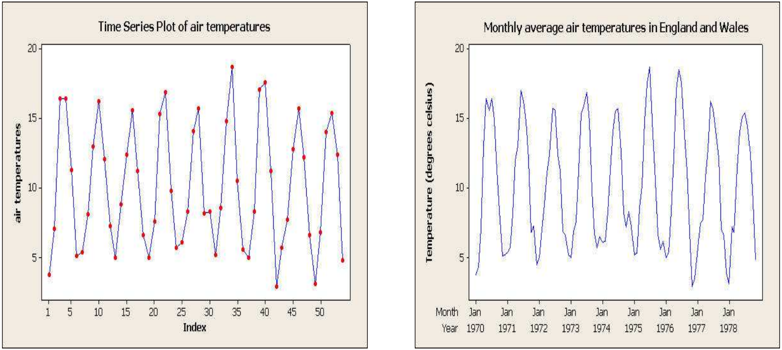

Figure: Left: Time series plots of monthly average air temperatures in England and Wales (right–hand–side plot shows edited time axis and inserted title)

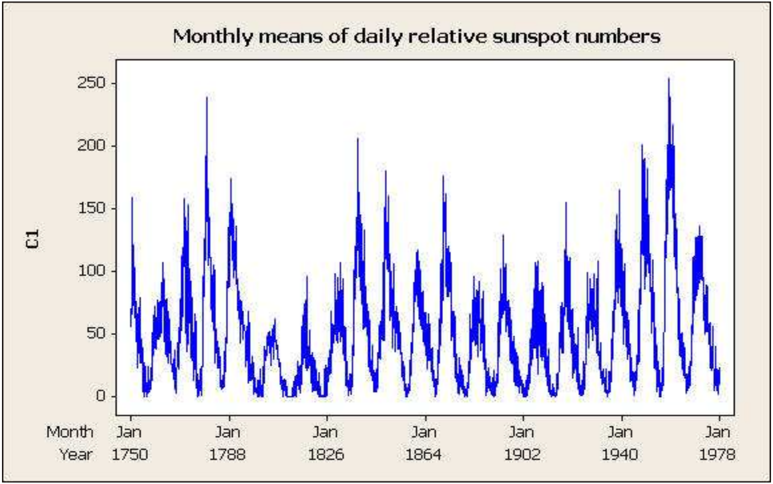

Figure: Time series plot of the monthly means of daily relative sunspot numbers

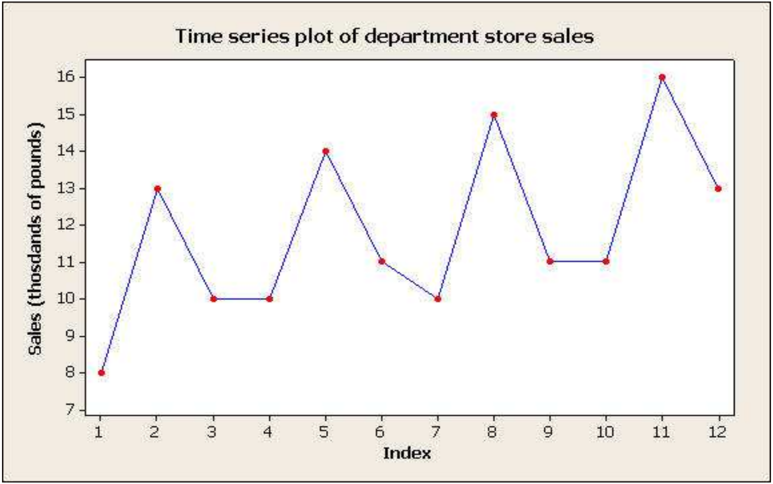

Figure: Time series plot of four–monthly sales for a department store

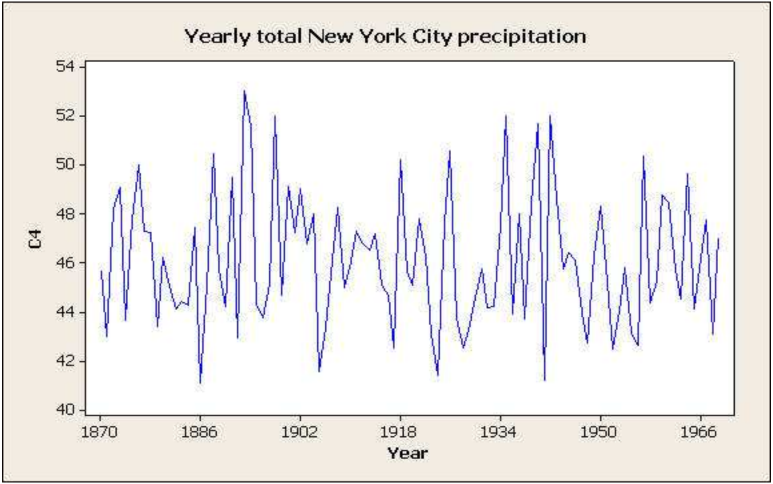

Figure: Time series plot yearly total precipitation in New York

Example

We will focus on the department store sales example. This one's easy to analyse by hand because it's relatively small!

The raw data are shown in the table below.

	Jan–Apr	May–Aug	Sep–Dec
1994	8	13	10
1995	10	14	11
1996	10	15	11
1997	11	16	13

Isolating the trend

There are lots of ways we can analyse the trend in this series. The simplest is to look at **moving averages**.

- To calculate moving averages, you first have to decide on the **cycle length**. This is the number of “seasons” you have, and is usually pretty obvious.
- In our example there are **3** seasons (Jan–Apr, May–Aug and Sep–Dec).

- The first moving average is just the mean of the first 3 observations, i.e.

$$\frac{8 + \mathbf{13} + 10}{3} \quad = \quad 10.33.$$

This moving average is centred around the second observation, or **time point 2**.

- The second moving average is just the mean of the next 3 observations, i.e.

$$\frac{13 + \mathbf{10} + 10}{3} = 11.$$

This moving average is centred around **time point 3**.

We proceed in this way to calculate all the moving averages. Note that there's no moving average for the first time point or the last time point.

The calculated moving averages are shown in the table below. A plot of these against time, along with the original data, is shown in the figure on the next slide.

	Moving averages		
	Jan–Apr	**May–Aug**	**Sep–Dec**
1994	*	10.33	11.00
1995	11.33	11.67	11.67
1996	12.00	12.00	12.33
1997	12.67	13.33	*

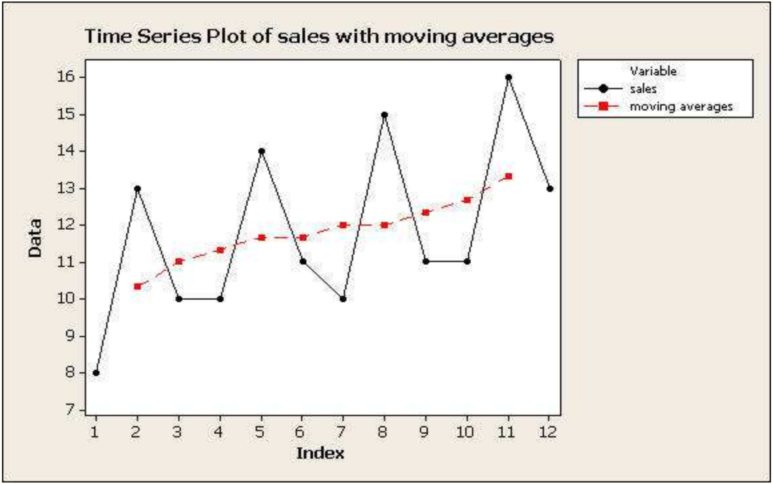

Figure: Time series plot of sales, with overlaid moving averages

Using regression to estimate the trend

Notice the linearity in the time series plot of moving averages. We can use **linear regression** to fit a line of best fit through the points!

Recall from lecture 6 that the simple linear regression model is

$$Y = \alpha + \beta X + \epsilon,$$

where

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \qquad \text{and} \\ \hat{\beta} &= \frac{S_{XY}}{S_{XX}} \end{aligned}$$

We now reformulate this model so that we have

$$Y = \alpha + \beta T + \epsilon,$$

where T is the "**time point**".

The first observation in the time series has a time point of 1 (i.e. $T = 1$).

The second observation in the time series has a time point of 2 (i.e. $T = 2$).

So if we wanted to predict sales in the period Jan–Apr 1998, we would substitute $T = 13$ into the regression equation above, since the last observed time point was for Sep–Dec 1997 and was the 12th time point.

But, before we do this, we need to estimate α and β!

The formulae for α and β are exactly the same as in chapter 6, but now X is replaced with T, i.e.

$$\begin{aligned} \hat{\beta} &= \frac{S_{TY}}{S_{TT}} \qquad \text{and} \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{t}, \end{aligned}$$

where

$$\begin{aligned} S_{TY} &= \left(\sum ty\right) - n\bar{t}\bar{y} \qquad \text{and} \\ S_{TT} &= \left(\sum t^2\right) - n\bar{t}^2. \end{aligned}$$

Remember, the easiest way to calculate these quantities is to draw up a table!

Note that for y we use the moving average values, *not* the actual observations!

Notice also that we don't have any observations at time points 1 and 12, since we were unable to calculate moving averages here.

t	y (moving averages)	ty	t^2
2	10.33	20.66	4
3	11.00	33.00	9
4	11.33	45.32	16
5	11.67	58.35	25
6	11.67	70.02	36
7	12.00	84.00	49
8	12.00	96.00	64
9	12.33	110.97	81
10	12.67	126.7	100
11	13.33	146.63	121
65	**118.33**	**791.65**	**505**

Thus,

$$\begin{aligned}\bar{t} &= \frac{65}{10} \\ &= \mathbf{6.5} \qquad \text{and}\end{aligned}$$

$$\begin{aligned}\bar{y} &= \frac{118.33}{10} \\ &= \mathbf{11.83}.\end{aligned}$$

Similarly,

$$\begin{aligned}S_{TY} &= \left(\sum ty\right) - n\bar{t}\bar{y} \\ &= 791.65 - 10 \times 6.5 \times 11.83 \\ &= \mathbf{22.7} \qquad \text{and}\end{aligned}$$

$$\begin{aligned}S_{TT} &= \left(\sum t^2\right) - n\bar{t}^2 \\ &= 505 - 10 \times 6.5 \times 6.5 \\ &= \mathbf{82.5}.\end{aligned}$$

Thus,

$$\begin{aligned}\hat{\beta} &= \frac{S_{TY}}{S_{TT}} \\ &= \frac{22.7}{82.5} \\ &= \mathbf{0.275} \qquad \text{and} \\ \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{t} \\ &= 11.83 - 0.275 \times 6.5 \\ &= \mathbf{10.043}.\end{aligned}$$

So, the regression equation for our trend is

$$\mathbf{Y} = \mathbf{10.043 + 0.275T} + \epsilon.$$

Isolating the seasonal effects

The linear trend we have estimated can be used to make forecasts, assuming our sales increase in a linear fashion.

Looking at our time series, we see that there are clear cycles around this linear increase.

We now look at how to **estimate these cycles**, known as **seasonal effects**.

Calculating seasonal effects

First of all, we calculate the **seasonal deviations** by subtracting the moving average for each observation from the original observation.

For example, the seasonal deviation for May–Aug 1994 is found as

$$13 - 10.33 \quad = \quad 2.67.$$

Similarly, for Sep–Dec 1994, we have

$$10.00 - 11.00 \quad = \quad -1.$$

The other seasonal deviations, along with the **seasonal means**, are shown in the table below:

	seasonal deviations		
	Jan–Apr	May–Aug	Sep–Dec
1994	*	2.67	–1
1995	–1.33	2.33	–0.67
1996	–2	3	–1.33
1997	–1.67	2.67	*
means	–1.67	2.6675	–1

We can now calculate the **seasonal effects**.

The seasonal effect for each season is the seasonal mean for that season minus the overall mean. The overall mean from the table above is found as

$$\begin{aligned} \frac{2.67 - 1 - 1.33 + \ldots - 1}{10} &= \frac{2.67}{10} \\ &= 0.267. \end{aligned}$$

Thus, the seasonal effects for each season are

$$\begin{aligned}\hat{s}_1 &= -1.67 - 0.267 \\ &= \mathbf{-1.937}\end{aligned}$$

$$\begin{aligned}\hat{s}_2 &= 2.6675 - 0.267 \\ &= \mathbf{2.4005} \qquad \text{and}\end{aligned}$$

$$\begin{aligned}\hat{s}_3 &= -1 - 0.267 \\ &= \mathbf{-1.267}.\end{aligned}$$

Note that the seasonal effects should add up to give zero. Ours don't – we have

$$\begin{aligned}\hat{s}_1 + \hat{s}_2 + \hat{s}_3 &= -1.937 + 2.4005 - 1.267 \\ &= -0.8035.\end{aligned}$$

Thus, we have to make an adjustment so they *do* add up to give zero.

To do this, we find the mean of our seasonal effects, and then subtract this from each of the seasonal effects. In this example, the mean of our seasonal effects is

$$\begin{aligned}\frac{-1.937 + 2.4005 - 1.267}{3} &= \frac{-0.8035}{3} \\ &= -0.26783.\end{aligned}$$

Thus, if we subtract this from each of the seasonal effects, the **adjusted** seasonal effects will then add up to give zero. Thus, the **adjusted** seasonal effects are

$$\begin{aligned}\hat{s}_1 &= -1.937 - (-0.26783) \\ &= \mathbf{-1.66917}\end{aligned}$$

$$\begin{aligned}\hat{s}_2 &= 2.4005 - (-0.26783) \\ &= \mathbf{2.66833}\end{aligned}$$

$$\begin{aligned}\hat{s}_3 &= -1.267 - (-0.26783) \\ &= \mathbf{-0.99917}.\end{aligned}$$

Just be careful with double negatives! Now the seasonal effects *do* sum to give zero!

Forecasting

There are many ways in which we can forecast future observations.

One way is to use the linear regression equation for the trend in our series. For the department store sales data, recall that this was

$$Y = 10.043 + 0.275T + \epsilon.$$

To predict average sales in Jan–Apr 1998, we would substitute $T = 13$ into the above equation, since this would be our 13th observation. Doing so, gives

$$\begin{aligned} Y &= 10.043 + 0.275 \times 13 \\ &= 10.043 + 3.575 \\ &= 13.618. \end{aligned}$$

However, we're not quite done yet! This assumes that our data follow a straight line!

Looking at the time series plot, clear cycles around an increasing trend can be seen.

Thus, we now need to **add** in our seasonal effect.

The seasonal effect for Jan–Apr is $\hat{s}_1 = -1.66917$. Thus, our "full" forecast for sales in Jan–Apr 1998 is

$$13.618 + (-1.66917) \quad = \quad 11.949,$$

i.e. £11,949, or just under £12,000.