

CHAPTER 14

of *Chance Encounters* by C.J.Wild and G.A.F. Seber

Time Series

14.1 Introduction

14.1.1 Measurements over time

In Chapter 13 we considered measurements over time of a special kind, namely those arising in a control chart. When the process is under control, the points form a sequence in time but with the property that all the points are identically distributed and independent, that is the points represent a random sample from a stable distribution. Although we can call such a sequence a time series, we usually reserve the term “time series” to describe a more general sequence in which the points are not necessarily independent and the distribution is not necessarily stable. It is always helpful to join up consecutive points in time. In Fig. 14.1.1a we have the scatter plot of a time series which is not at all informative until we join up the points as in Fig. 14.1.1b.

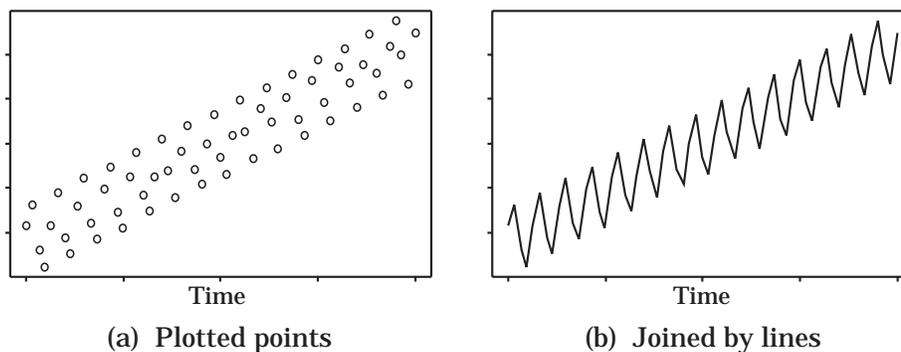


Figure 14.1.1 : The same time series graphed 2 ways.

Example 14.1.1 Over recent years the Climatic Research Unit at the University of East Anglia, England, has been compiling and checking temperature records from over 3000 land-based sites and using nearly 80 million observations of marine temperatures taken by both the merchant and naval fleets of

2 Time Series

the major maritime nations. After a great deal of sifting and sorting of the data, they have come up with a time series of the mean global temperatures from 1854 to 1994. Because the actual changes are small they have subtracted off the average for the years 1950–1979: this accentuates the changes. This time series is plotted in Fig. 14.1.2 and the picture suggests that we have an overall trend overlaid with a random looking irregular component. Taking a line by eye through the “middle” of the data it appears that the “average” or smoothed temperature (i) is approximately constant from 1845 to about 1920, (ii) steadily rises from 1920 to about 1940, (iii) remains roughly constant or perhaps drops slightly from about 1940 to about 1975 and, finally, (iv) rises steadily in an almost linear fashion from 1975 to 1994. There is a clear indication that the average temperature of the world has risen, in fact, by about 0.6°C over the past 140 years. It is also clear that we can expect a lot of annual fluctuations in the future. Summing up, it appears that this time series could be decomposed into two components, a trend plus an irregular or “random” component. If we wanted to predict the temperature in the future we could begin by fitting a trend model to the data, then subtracting off the trend and fitting some model to the irregular component that is left. To predict we could then extrapolate both models into the future and add the two results together.

■

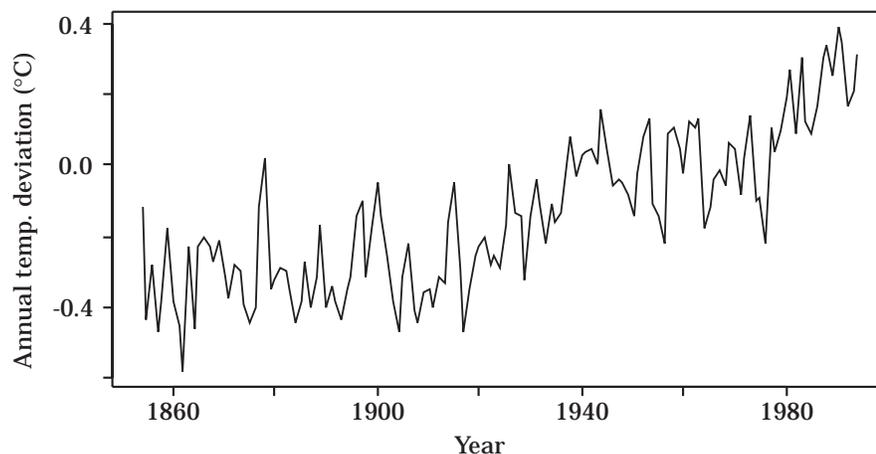


Figure 14.1.2: Deviation in the average global temperature (in $^{\circ}\text{C}$).
[Data courtesy of Dr Philip Jones, Climate Research Unit, Univ. of East Anglia.]

A time series we might see in a newspaper or magazine, is given in Fig. 14.1.3 describing how some index or measurement, in this case percentage full-time unemployment, varies over time. Typically the time interval may be months, three months (that is, quarterly as in Fig. 14.1.3) or years. Apart from the fluctuations, one would hope that the series will give some indication of any underlying trend that may be present. We can imagine that what happens in the near future will depend very much on what is happening now. For example,

if the unemployment rate is high this quarter it will tend to be high the next quarter, as in Fig. 14.1.3 . This is an example of positive association or, more technically, **positive correlation** between observations that are next to each other. On the other hand, we might envisage a process in which a high value at one point in time leads to over-compensation so that there is low value next time and the sequence tends to oscillate backwards and forwards. This is an example of **negative correlation**. If we could model this correlation effect in some way we could then use the model to help us make predictions.

The time series in Fig. 14.1.3 is plotted from quarterly data¹ (March, June, September and December) for NZ. We note that it is fairly smooth and shows a general downward trend. However other time series, like the one in Fig. 14.1.4 showing monthly rainfall data collected near Auckland University from 1949 to 1994 inclusive, can be very “up and down” with apparently little or no evidence of a trend.

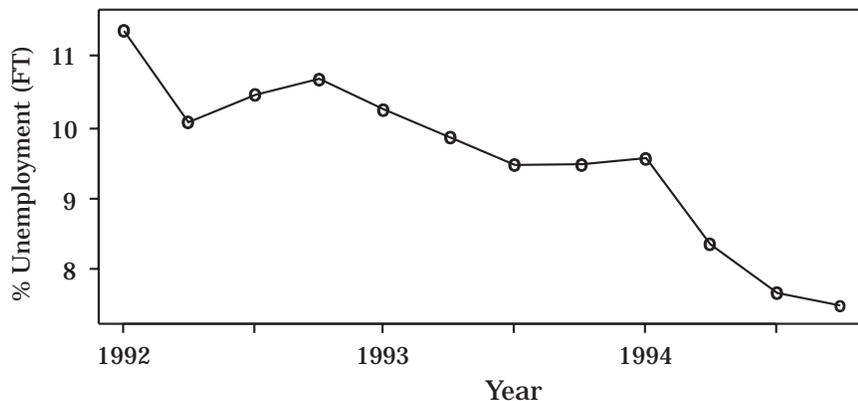


Figure 14.1.3 : Quarterly data for % full-time unemployment.²

As in weather data, we can also expect seasonal (quarterly) variation in many economic time series. For example part-time employment will vary with the time of the year as seasonal work such as fruit picking etc. becomes available. In Fig. 14.1.5 (plotted from the data in Table 14.2.2, which is used later) we have a series like Fig. 14.1.3 but now with part-time³ instead of full-time unemployment. We note that the percentage goes up for the first

¹There is some ambiguity in plotting quarterly data. We can either have the first quarter plotted at the beginning of each year, as we have done, or we can shift the whole graph along one quarter so that the fourth quarter corresponds with the beginning of the next year.

²For unemployed people, full-time means seeking employment for 30 or more hours per week.

³Unemployed people requiring less than 30 hours employment per week.

4 Time Series

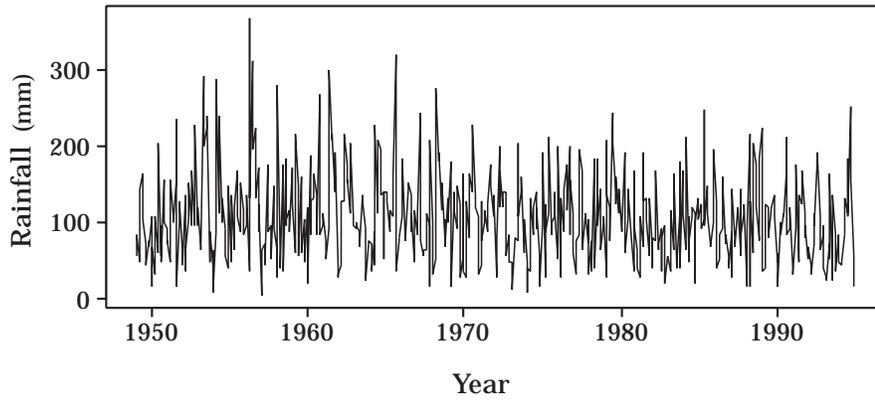


Figure 14.1.4 : Monthly rainfall data for Auckland.
[By courtesy of Dr C. M. Triggs, Univ. of Auckland.]

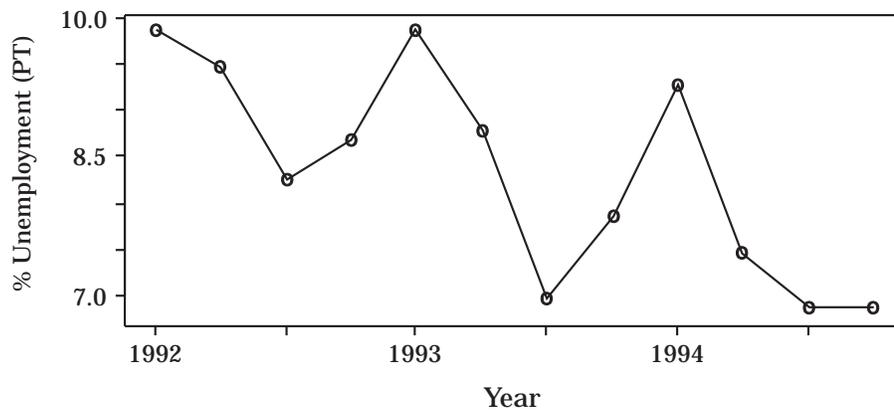


Figure 14.1.5 : Quarterly data for % part-time unemployment.

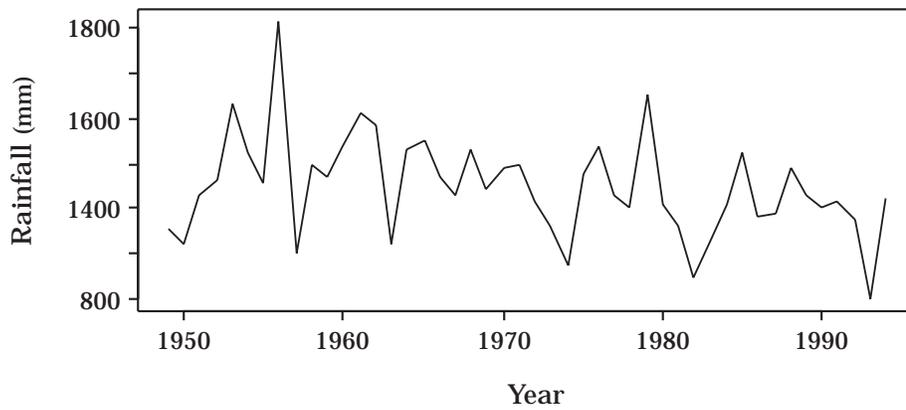


Figure 14.1.6 : Annual rainfall data for Auckland.

quarter of each year (March⁴) but drops to a low in the third quarter (September) at the end of the New Zealand winter. In this particular example we have a combination of a trend, a quarterly cycle and irregular variation.

Suppose we wanted to predict what would happen in the future. If we are only interested in annual predictions, then we could simply use annual totals obtained by adding the data for the four quarters in each year, thus avoiding the problem of seasonal variation. However, if we want to predict just one quarter ahead then we would need to take the quarterly variation into account. Similar comments apply to the monthly rainfall data. If we are only interested in annual predictions then we could pool the monthly data and plot just the annual totals, as in Fig. 14.1.6. Now that we have got rid of some of the “noise” caused by all the monthly variation there seems to be a slow downward trend indicating perhaps a long-term fluctuation in the weather.

Such long fluctuations mentioned above are not uncommon and are often referred to as “cycles”. The word cycle is a bit misleading as it is usually reserved in scientific circles to mean periodic phenomena like a sine curve. Quarterly variation belongs to this periodic category. However, these long fluctuations may not be periodic or they may be “periodic” with a variable period and are therefore difficult to predict. So-called “business cycles” come into this category and their existence is somewhat controversial. Instead of treating a long-term fluctuation as some form of cycle we can split up the series into sections and fit a separate trend to each section: so-called “piecewise” fitting.

Clearly there are a number of different factors that can affect a time series. To make predictions we need to find and exploit any regular features that can be found in the data like trends, cycles such as seasonal variation, and residual correlation between data points. All of these have been demonstrated in previous examples. In Fig. 14.1.7 we show how various components may add together to get the final series. If we wanted to predict a future observation we could try and model all three components, extrapolate the models to the appropriate future time and then combine the results to get our overall prediction.

It should be noted that data only gives us predictive information if we can assume that the patterns we see in past data will continue into the future. This assumes some stability in the mechanism producing the data. Unexpected changes in these mechanisms obviously falsify the forecast. We therefore need to watch out for such things as a spike (e.g. Fig. 14.1.8a) which signals an outlier, or a step (e.g. Fig. 14.1.8b) which signals a sudden change like a devaluation in the currency.

As the subject of time series is a big one, we can only give a taste in this chapter! We will focus on some basic ideas rather than on a lot of techniques.

⁴This is end of the New Zealand summer.

6 Time Series

A related problem, which we shall not consider, is signal processing where we endeavor to detect the presence of a signal in amongst the random noise.

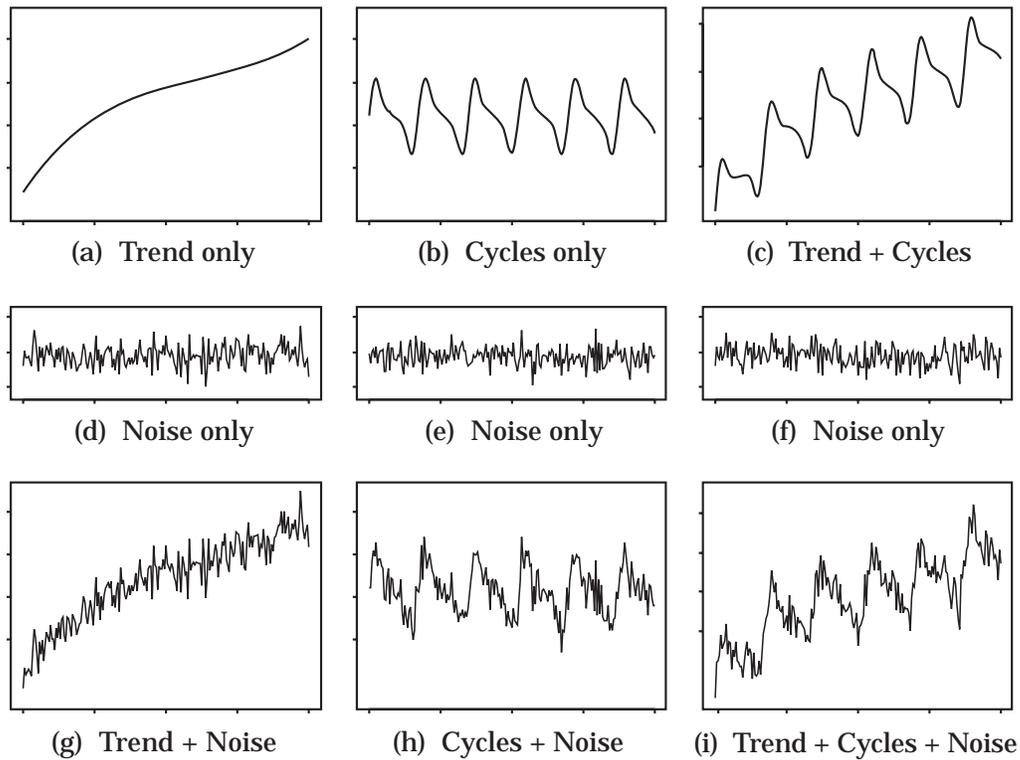


Figure 14.1.7: Combining trend, cycles and noise.

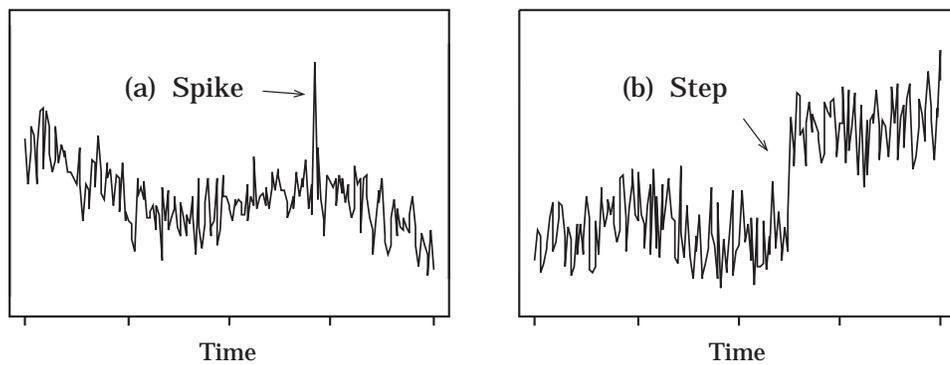


Figure 14.1.8: Features to look out for: (a) spikes and (b) steps.

Exercises on Section 14.1.1

1. In Example 14.1.1 we discussed some data on average global temperature deviations. The plot of the data in Fig. 14.1.2 suggested that there might be a linear trend in the temperature over the past 20 years. To check this more closely, the 20 years of data are given in Table 14.1.1 below. Plot the time series and comment on any trends or other features of the data.

Table 14.1.1 : Deviation (y_t) in the Average Global Temperature in $^{\circ}C$

Year	y_t	Year	y_t	Year	y_t	Year	y_t
1975	-0.09	1980	0.19	1985	0.09	1990	0.39
1976	-0.22	1981	0.27	1986	0.17	1991	0.35
1977	0.11	1982	0.09	1987	0.30	1992	0.17
1978	0.04	1983	0.30	1988	0.34	1993	0.21
1979	0.10	1984	0.12	1989	0.25	1994	0.31

Source: Courtesy of Phil Jones, Climate Research Unit, University of East Anglia.

2. In Table 14.1.2 we have the annual consumption of tobacco products, measured in units of “cigarette equivalents”, in the US from 1960–1986. Plot the time series and comment on any features. What do you conclude from the plot?

Table 14.1.2 : US Tobacco Consumption in Cigarette Equivalents per Adult per Year^a

Year	y_t	Year	y_t	Year	y_t	Year	y_t
1960	4655	1967	4613	1974	4263	1981	3830
1961	4670	1968	4506	1975	4167	1982	3732
1962	4626	1969	4322	1976	4104	1983	3502
1963	4681	1970	4300	1977	4028	1984	3443
1964	4700	1971	4310	1978	3943	1985	3358
1965	4672	1972	4282	1979	3880	1986	3275
1966	4660	1973	4311	1980	3837		

^aSource: Statistics New Zealand.

3. In Table 14.1.3 we have annual information on cigarette consumption versus cigarette price in New Zealand. The prices are converted to the equivalent price in December 1990 so as to allow for the effects of inflation.
 - (a) Draw time series plots for each of the variables “Cigarette consumption” and “Cigarette price”. What do you conclude?
 - (b) Plot consumption versus price. Does this show that the price increases have driven down consumption? What other conclusions are possible?

Table 14.1.3 : Consumption and Price of Cigarettes^a

Year	Cigarette consumption Number per adult ^b	Price per packet ^c (1990)\$
1973	2691	3.05
1974	2732	2.78
1975	2874	2.56
1976	2827	2.51
1977	2844	2.54
1978	2783	2.41
1979	2706	2.41
1980	2617	2.42
1981	2666	2.39
1982	2602	2.48
1983	2541	2.63
1984	2560	2.58
1985	2293	2.67
1986	2102	3.04
1987	2125	3.41
1988	2101	3.48
1989	1661	4.12
1990	1725	4.30
1991	1541	4.75

^aSource: Statistics New Zealand, 1994. ^bAged 15 years and over.

^cPer packet of 20, inflation adjusted to 1990 price-equivalent.

14.1.2 Regression and Time Series

Time series models can be developed in a somewhat similar fashion to regression models. The observation y_t at time t ($t = 1, 2, \dots, T$) is regarded as a sum of two components, namely,

$$y_t = \text{fixed component} + \text{random error},$$

where the fixed component in regression was called trend. This model is called an *additive* model as the “fixed component” and the “random error” add together. Another possibility is the *multiplicative* model

$$y_t = \text{fixed component} \times \text{random error}.$$

This can be changed into an additive model by taking logarithms.⁵ However, we saw above that a time series may have not only a trend but it may also have a cyclic or seasonal component in the fixed part of the model. In contrast to the regression model where we assume the random errors to be a random sample from a distribution, we now allow the errors in a time series to no longer be independently and identically distributed but to be correlated in some way. Analyzing a time series therefore requires some detective work to

⁵ $z_t = \log y_t = \log(\text{fixed component}) + \log(\text{random error})$.

find out what components are in the fixed part of the model and what sort of correlation structure is in the random or error part. As we mentioned before, both the fixed components and the correlation structure can be used for making predictions. To help us look for trends we now look at some smoothing methods.

Quiz for Section 14.1

1. What is the difference between a control chart and a time series as discussed above?
2. Describe the difference between positive correlation and negative correlation for adjacent observations in a time series. What would the time series plot look like in each case?
3. What ambiguity must be resolved before plotting monthly data?
4. What three components should one look for in analyzing a time series?
5. What is the difference between an additive time series model and a multiplicative one?

14.2 Smoothing Out the Bumps

14.2.1 Moving averages

If we are going to use a time series for prediction we need to find any trend that might exist in the data. As trends tend to be obscured by the random errors, some smoothing method is needed to iron out some of the “ups” and “downs”.

The simplest way of smoothing a time series is to use a moving average m_t which is based on averaging adjacent time periods.⁶ If we average the observations in pairs then we average observations one and two, two and three, three and four etc. so that the average steadily moves along the time period. We effectively replace each consecutive pair by a smoothed value located midway between the two points so that the smoothed series is out of phase with the original series.

With a moving average based on just two observations we may still get a lot of fluctuations and not enough smoothing. If we average the observations y_t three at a time, i.e. average observations one, two and three, then observations two, three and four etc., we will get a smoother curve.⁷ The more points we use to calculate the average, the smoother the curve we will get. In general, a k -point moving average is one based on averaging k adjacent time periods. How big should k be? In order to decide this we notice first of all that we “lose” some points at the beginning and end of the smoothed series. For example if $k = 3$, then the first three observations will give the first average and this average will be located at the middle of the interval, namely the second time period. Using the same reasoning, the last moving average will be located at

⁶This is often referred to as a filtering process.

⁷Instead of using the mean we could use the median, that is the middle of the three observations, though it can lead to anomalies. However it is a useful, robust method particularly if the procedure is carried out several times, that is, one re-smooths the smoothed series. See Tukey [1977, Chapter 7].

the second to last time period. Thus, in effect, each point is replaced by the average of this point with the points before and after. This means that we have no smoothed points for the first and last time periods. If we average 5 points then we are short of two smoothed points at each end. We can, of course, use the original points at the ends or extend the smoothed curve at each end by eye.⁸ When k is an odd number, the smoothed series is now “in phase” with the original series.

3-point moving average smoothing: Replace each observation by the mean of the original observation and the observation on either side.

Table 14.2.1 : Annual Percentage Increase in CPI for the US and 3-point Moving Average^a

Year	% increase in CPI (y_t)	3-point moving total	3-point moving average (m_t)
1985	4.1	–	–
1986	3.5	9.2	3.07
1987	1.6	9.2	3.07
1988	4.1	11.0	3.67
1989	5.3	14.3	4.77
1990	4.9	15.7	5.23
1991	5.5	14.0	4.67
1992	3.6	12.2	4.07
1993	3.1	9.5	3.17
1994	2.8	–	–

^aSource: Statistics New Zealand, 1994.

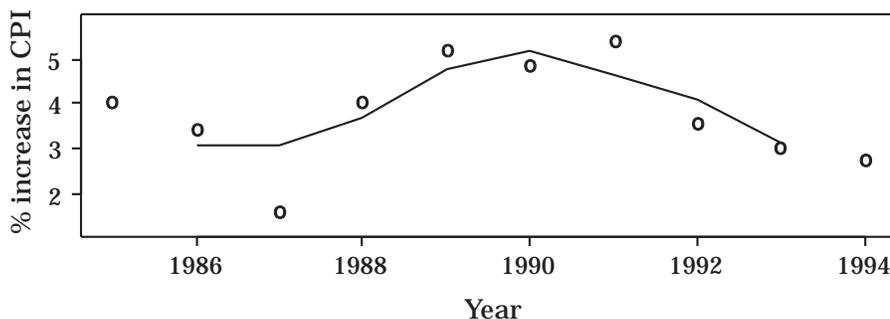


Figure 14.2.1 : Annual percentage increase in the CPI for the US (points) and a 3-point moving average (lines).

⁸To do this more rigorously we would have to fit some sort of model to the data. For example we could fit a polynomial (curve) to the last few smoothed values.

Example 14.2.1 Because inflation is always with us we find that the CPI or Consumer Price Index, which is often used as a measure of inflation (see Section 14.4) invariably increases over time. What is of particular interest is the rate of increase. In Table 14.2.1 we have a time series representing the percentage increase in the CPI from the previous year for the years 1985 to 1994 in the United States. The CPI used is an annual average from one March to the next. The original series (shown as points) and the smoothed series (shown as lines) are graphed in Fig. 14.2.1. ■

With the above 3-point smoothing we saw that the end points are not smoothed. Also the smoothing uses only three points and ignores the rest of the time series. There is another form of smoothing called exponential smoothing which does take into account the rest of the series. We shall not use it as it should only be applied to a time series with no trend or cyclical variation (a so-called “stationary” time series). However it would be a useful tool for smoothing the irregular component of a time series after the trend and cycle have been removed.⁹

Exercises on Section 14.2.1

- Using the data in Table 14.2.1 plot the 3-point moving median series and compare it with the 3-point moving average plot. The moving median is obtained by replacing each observation by the median of the original observation and each observation on either side. How does your plot compare with Fig. 14.2.1?
- Referring to Fig. 14.1.2 in Section 14.1.1, there does not seem to be any overall temperature trend from 1860 to 1920. To look at this more closely we consider just the series of data from 1860 to 1889 given below.

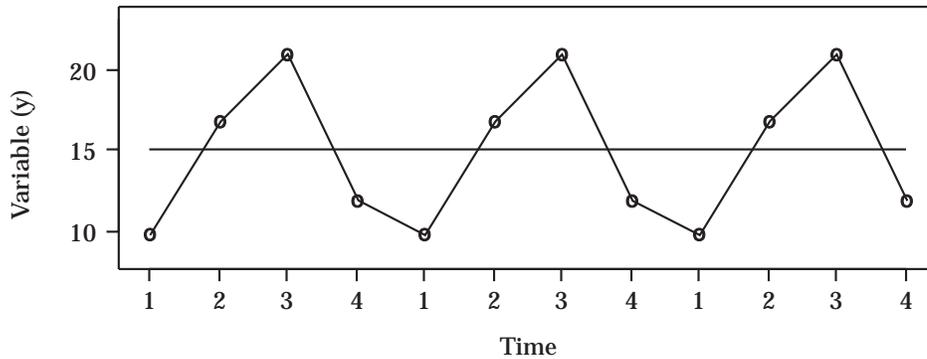
Year	x_t	Year	x_t	Year	x_t
1860	-0.38	1870	-0.31	1880	-0.32
1861	-0.45	1871	-0.37	1881	-0.29
1862	-0.58	1872	-0.28	1882	-0.30
1863	-0.23	1873	-0.30	1883	-0.35
1864	-0.46	1874	-0.39	1884	-0.44
1865	-0.23	1875	-0.44	1885	-0.38
1866	-0.20	1876	-0.40	1886	-0.27
1867	-0.23	1877	-0.12	1887	-0.40
1868	-0.27	1878	0.02	1888	-0.31
1869	-0.21	1879	-0.35	1889	-0.17

- (a) Plot the series. (If you are doing it by hand you may find it helpful to add 1 to each observation to simplify the calculations.)

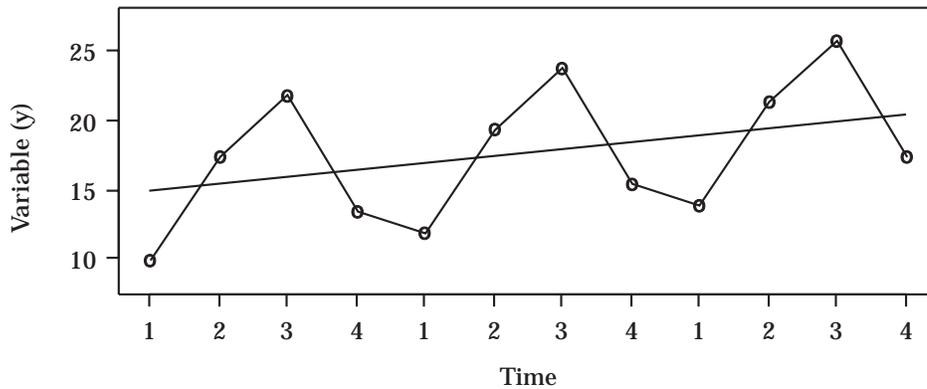
⁹It also forms the basis of the so-called Holt-Winters method which can be applied to a time series which contains both trend and seasonal variation. See Chatfield [1996] for references.

12 Time Series

- (b) On the same graph draw in the 3-point moving average. Do you think that a 5-point moving average might be more helpful? Try it and see.
 - (c) Is there a linear trend? Do you think that a moving average is good at detecting a linear trend?
3. One method for removing a linear trend is by differencing, that is by plotting the series with terms $z_t = y_{t+1} - y_t$. Verify that if this differencing is applied to the straight line $y_t = a + bt$ then z_t does not contain a trend.



(a) With no trend



(b) With linear trend

Figure 14.2.2 : Perfect quarterly variation.

14.2.2 Smoothing quarterly variation

In Fig. 14.2.2(a) we have an idealized economic time series in which there is variation from quarter to quarter but the pattern of variation is exactly the same every year. What sort of moving average would smooth this? The data for the first 12 quarters of the series are as follows:

10, 17, 21, 12; 10, 17, 21, 12; 10, 17, 21, 12.

Here a semicolon separates each year. If we take *any* four consecutive observations from this series we see that we get the same four numbers, but in a

different order (try it!). Therefore using a moving average of four observations will give a constant mean. This is represented by the horizontal straight line in Fig. 14.2.2(a).

Suppose we now take this figure and rotate it slightly as in Fig. 14.2.2(b). Then a 4-point smoothing is still given by a straight line but now sloping upwards indicating an increasing trend. The data for this series is:

10, 17.5, 22, 13.5; 12, 19.5, 24, 15.5; 14, 21.5, 26, 17.5.

In both cases a 4-point moving average has smoothed out the seasonal effect and has indicated the presence or absence of any trend. Let's now try this 4-point smoothing on some real data.

Example 14.2.2 In Table 14.2.2 we have data representing the percentage part-time unemployment in New Zealand. This was originally plotted in Fig. 14.1.5. When we use a moving average we place the smoothed value at the center of the range of points used. For the first four points this corresponds to a time midway between the second and third points. Doing this for consecutive sets of four we get the column labeled $(y'_{t+.5})$ in Table 14.2.2. To get this smoothed series back in step with the original series we now average consecutive pairs, giving the column labeled (z_t) .¹⁰ For example

Table 14.2.2 : Part-time Unemployment Rates with Centered 4-point Smoothing^a

Quarter	% (y_t)	4-point smoothing ($y'_{t+.5}$)	2-point smoothing of y'_t (z_t)
1992 Mar	9.9		
Jun	9.5		
Sep	8.3	9.100	9.100
Dec	8.7	9.100	9.0125
1993 Mar	9.9	8.925	8.7625
Jun	8.8	8.600	8.5000
Sep	7.0	8.400	8.3250
Dec	7.9	8.250	8.0875
1994 Mar	9.3	7.925	7.9125
Jun	7.5	7.900	7.7750
Sep	6.9	7.650	
Dec	6.9		

^aSource: Statistics New Zealand.

¹⁰These two steps may be combined into a single step as follows: To find the smoothed value at time t , take y_t , add to it the values of its neighbors (y_{t-1} and y_{t+1}), then add in half the values of the next neighbors (y_{t-2} and y_{t+2}), and divide by 4, namely

$$z_t = \frac{1}{4}(\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}).$$

14 Time Series

$$y'_{2.5} = \frac{1}{4}(y_1 + y_2 + y_3 + y_4) = \frac{1}{4}(9.9 + 9.5 + 8.3 + 8.7) = 9.100,$$

$$y'_{3.5} = \frac{1}{4}(y_2 + y_3 + y_4 + y_5) = \frac{1}{4}(9.5 + 8.3 + 8.7 + 9.9) = 9.100$$

and

$$z_3 = \frac{1}{2}(y'_{2.5} + y'_{3.5}) = \frac{1}{2}(9.100 + 9.100) = 9.100.$$

The smoothed series z_t is plotted with lines in Fig. 14.2.3 along with the individual points of the original series y_t . (You should check some of the values of z_t for yourself.) The fact that z_t is approximately linear indicates that the seasonal and random variation have been largely smoothed out. We note that z_t has two points missing from each end.¹¹

Finally we note that the above method can be used for smoothing out any cycle which repeats itself every c months or years. We simply take a c -point moving average. The key step is to determine the period c of the cycle.

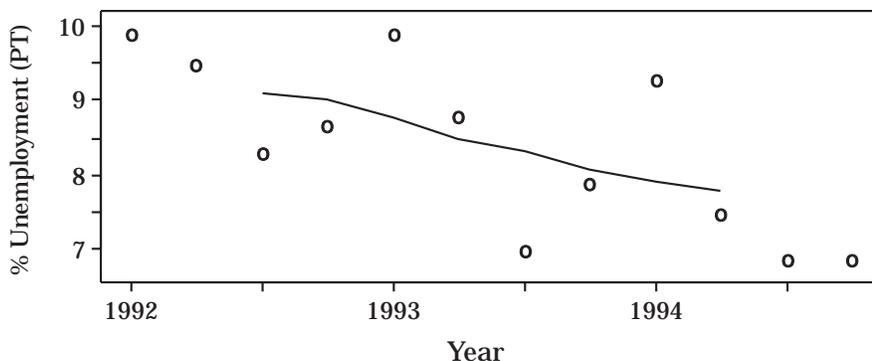


Figure 14.2.3: Centered 4-point moving average for unemployment data.

Exercise on Section 14.2.2

1. In Table 14.2.3 we have the quarterly numbers of people, in thousands, arriving in New Zealand. Plot the series and a 4-point centered moving average. Is there any evidence of a trend?

Table 14.2.3: Quarterly Arrivals in New Zealand (in '000)^a

1990 Mar	239	1991 Mar	238	1992 Mar	249	1993 Mar	274
Jun	204	Jun	193	Jun	202	Jun	220
Sep	238	Sep	243	Sep	243	Sep	260
Dec	253	Dec	262	Dec	274	Dec	295

^aSource: Statistics New Zealand.

¹¹We have ignored this end-point feature in plotting the seasonally adjusted series in Figs 14.2.2(a) and (b). However, with such perfect linear smoothing there, it would not be unreasonable to extend the straight lines in both directions, as we have done.

14.2.3 Deseasonalized time series

When we use 4-point smoothing we smooth out not only the seasonal variation but also the random (irregular) variation that is always present throughout a time series. What we might like to do is remove just the seasonal effect and leave any trend and the random ups and downs back in the data. The resulting series gives us what is known as “deseasonalized” data which may give us a clearer picture of what is happening. To achieve this in a sensible manner we need a suitable model for the process producing the original time series. We have already mentioned that two possible models are an additive model and a multiplicative model, the latter being converted into an additive model by taking logarithms. We therefore focus on just additive models using as our data either y_t or $\log y_t$. Our model takes the form

$$\text{data} = \text{trend} + \text{cycle} + \text{error}.$$

To remove a quarterly cycle, for example, we begin by averaging all the first quarters, namely $(y_1 + y_5 + y_9 + \dots)/\text{no. of years}$, then averaging all the second quarters, all the third quarters and finally all the fourth quarters, giving us just four numbers s_1, s_2, s_3 and s_4 . We then subtract the mean \bar{s} of these four numbers (which is the same as \bar{y} of the original series) to get $s_t - \bar{s}$. The deseasonalized series is then given by $z_t = y_t - (s_t - \bar{s})$, where the definition of s_t is extended beyond the first year by simply repeating the same four numbers.

How do we know when to use $\log y_t$ instead of y_t ? If we are interested in proportional or percentage changes then taking logarithms may be more sensible. If the “height” variation of the irregular component or cycle tends to increase as the trend increases then these components would appear to be having a multiplicative effect so that we would try logarithms. If we are not certain then we can try both y_t and $\log y_t$ and see which works best. We must experiment!

Example 14.2.2 cont. We refer once again to the y_t in Table 14.2.2. These are reproduced for ease of reference in Table 14.2.4.

To demonstrate the calculations, we have

$$s_1 = s_5 = s_9 = \frac{1}{3}(y_1 + y_5 + y_9) = \frac{1}{3}(9.9 + 9.9 + 9.3) = 9.7,$$

$$\bar{s} = 8.38,$$

$$s_1 - \bar{s} = 9.9 - 8.38 = 1.32$$

and

$$z_1 = y_1 - (s_1 - \bar{s}) = 8.58.$$

The series z_t is plotted as lines in Fig. 14.2.4 along with the points for y_t . As the method is meant to provide just a graphical aid rather formal estimates we have carried out the calculations to just two decimal places. Without the seasonal cycle there appears to be an almost linear trend.

Table 14.2.4 : Deseasonalized Part-time Unemployment Rates^a

Quarter	% (y_t)	Quarters (s_t)	Adjustment ($s_t - \bar{s}$)	Deseason. (z_t)
1992 Mar	9.9	9.7	1.32	8.58
Jun	9.5	8.6	0.22	9.28
Sep	8.3	7.4	-0.98	9.28
Dec	8.7	7.83	-0.55	9.25
1993 Mar	9.9	9.7	1.32	8.58
Jun	8.8	8.6	0.22	7.58
Sep	7.0	7.4	-0.98	7.98
Dec	7.9	7.83	-0.55	8.45
1994 Mar	9.3	9.7	1.32	7.98
Jun	7.5	8.6	0.22	7.28
Sep	6.9	7.4	-0.98	7.88
Dec	6.9	7.83	-0.55	7.45
Mean	8.38			

^aSource: Statistics New Zealand.

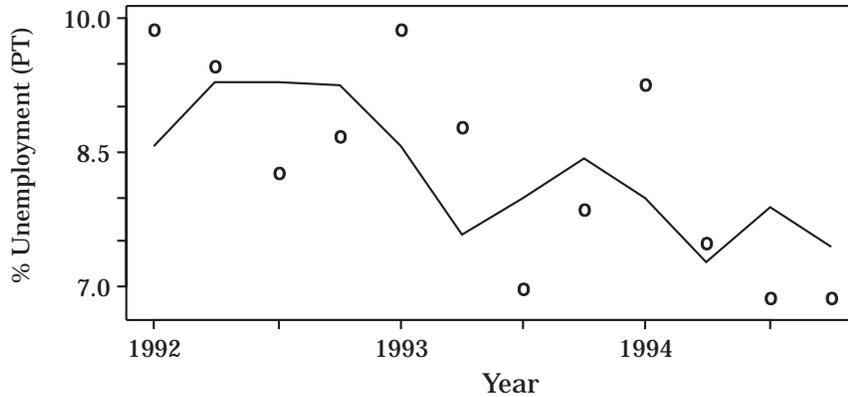


Figure 14.2.4 : Deseasonalized part-time unemployment rates.

Exercise on Section 14.2.3

Deseasonalize the data in Table 14.2.3. What do you conclude?

Quiz for Section 14.2

1. In what way does a smoothed time series based on a moving average of an odd number of consecutive values differ from one based on an even number of values?
2. Which moving average provides a smoother series, one based on 3 points or one based on 5 points?
3. Can you think of one advantage and one disadvantage that a 3-point moving median has over a 3-point moving average?
(Section 14.2.1)
4. How many points are used for smoothing quarterly variation? How many would you use for smoothing a monthly cycle?
5. How do we center 4-point smoothing?

6. What are the three components making up a time series?
 7. How can we convert a multiplicative model into an additive one?

(Section 14.2.2)

(Section 14.2.3)

*14.3 Forecasting¹²

The above methods of smoothing and deseasonalizing have been introduced as informal tools to help us look out for trends and other features of a time series. We may even be able to make a reasonable forecast by simply extending smoothed lines by eye into the future. However, when it comes to making a formal prediction the approach is different, which we now consider. We just give an outline. Our basic approach is to break the time series y_t into components, predict each component for some future time and then add the answers together.

Suppose we use moving-average smoothing and find that there is a clear trend. If the trend is roughly linear, we can fit a straight line to the scatter plot of the *original data* y_t versus t using the method of least squares described in Chapter 12.¹³ However an important assumption that we require for our least squares theory to hold is that the observations are independent. As we have already emphasized at the beginning of this chapter, we can generally expect adjacent observations to be correlated and therefore not independent. Fortunately all is not lost! It turns out that we can still use the least squares estimates of the slope and intercept as these remain unbiased estimates even in the presence of correlation. We also get valid residuals. What are affected by the correlations are the standard errors which we do not consider in this chapter.

If the plot is gently curving we can try fitting a quadratic equation or some other polynomial rather than a straight line. Alternatively, the plot can often be “straightened” by taking logarithms and then fitting a straight line to the $\log y_t$ data.¹⁴ The above comments about the validity of least squares fitting generally apply to these models as well. For some series we find that the trends are relatively short term and the data goes through a long-term fluctuation.¹⁵

Once we have fitted the trend, \hat{y}_t say, we can subtract it from the original series to get the residuals $r_t = y_t - \hat{y}_t$.¹⁶ The next step is to remove any cycle. With seasonal data, for example, we can compute the seasonal component $s_t - \bar{s}$ from r_t as in Section 14.2.3 and then subtract it from r_t to get x_t , say. We can

¹²This topic is introduced for completeness only.

¹³It should be noted that the moving average is not generally recommended by itself for measuring trend (Chatfield [1996]).

¹⁴This assumes that the trend is of the exponential form $z = ae^{bt}$. Then $\log_e z_t = \log_e a + bt = a_0 + bt$, which is a straight line.

¹⁵In this case we may be able to fit each short-term trend in a piecewise fashion.

¹⁶We can also remove a linear trend by “differencing” the original series to get $y_{t+1} - y_t$.

do this with any cycle. In fact there may be several cycles with different periods to be removed. Cycles can also be removed with differencing techniques. The series x_t will then reflect just the irregular or “error” component as the trend and any cycles have been removed. This remaining component can be modeled several ways. One approach is to choose a model from a suitable family of well-defined models which best fits the series x_t . One such family of models is the so-called family of ARIMA models.¹⁷ Unfortunately these require a statistical theory beyond the scope of this text.

To predict y_t for some time t_p in the future we could extend our fitted trend to time t_p then add on $(s_{t_p} - \bar{s})$ if there is a seasonal or cyclic component, and finally add on the predicted value of x_{t_p} . A cruder method would be to ignore the last step: this would give us a prediction without the error component.

An important time series that we hear a lot about is the so-called Consumer Price Index or CPI. Before we discuss it and find out how it is constructed we introduce the concept of an index number.

14.4 Index Numbers

In many subject areas such as Economics we are often concerned with relative (e.g. percentage) change rather than the absolute change. We can therefore use an **index number** which is simply a number that measures relative change in a time series variable y_t . To describe how an index number is computed, consider the time series given by column 2 in Table 14.4.1. Here y_t is the price of a particular appliance produced by a company from the years 1988 to 1994. We begin by choosing a base period such as 1988 and note the value of y_{1988} . To find the index number I_t at time t we compute the ratio y_t/y_{1988} and convert it to a percentage, namely

$$I_t = \frac{y_t}{y_{1988}} \times 100.$$

Table 14.4.1 : Annual Prices of an Appliance with Index Numbers for Two Base Years

Year	y_t Price(\$)	I_t base=1988	I_t base=1991
1988	586	100	94.5
1989	601	102.6	96.9
1990	632	107.8	101.9
1991	620	105.8	100
1992	645	110.1	104.0
1993	668	114.0	107.7
1994	699	119.3	112.7

¹⁷These were developed by Box and Jenkins and are described in various books on time series such as Chatfield [1996], who gives further references.

For example

$$I_{1990} = \frac{y_{1990}}{y_{1988}} \times 100 = \frac{632}{586} \times 100 = 107.8$$

and

$$I_{1994} = \frac{y_{1994}}{y_{1988}} \times 100 = \frac{699}{586} \times 100 = 119.3.$$

The index for the base period is

$$I_{1988} = \frac{y_{1988}}{y_{1988}} \times 100 = 100,$$

and this is always true for the base period.¹⁸ The rest of the index numbers are given in column 3 of Table 14.4.1 and are plotted in Fig. 14.4.1. This time series plot will give us an rough idea of *relative* change.

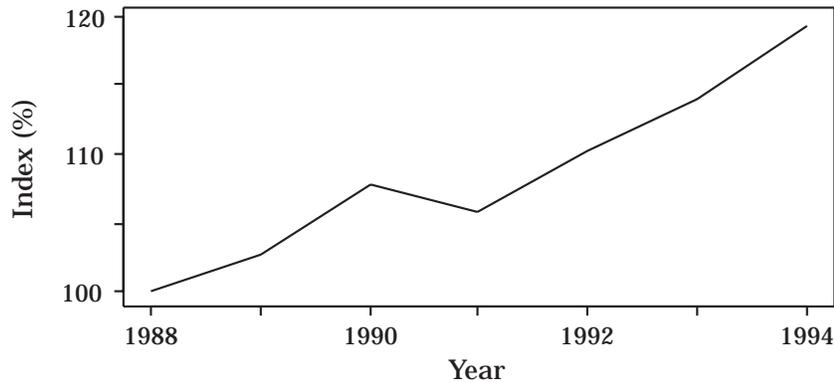


Figure 14.4.1 : Graph of the index number with base period 1988.

How can such an index be used? Since the index in 1990 is 107.8 we can say that the appliance costs 7.8% more in 1990 than it did in 1988. Also, in 1994 the index is 119.3 so that from 1990 to 1994 the price index has increased by $119.3 - 107.8 = 11.5$ percentage points. However it is not correct to say that there was an 11.5% increase from 1990 to 1994. The ratio of prices during that time is $\frac{119.3}{107.8} \times 100\%$ or 110.7% so that the correct increase over that period is 10.7%.¹⁹

The index number is
$$I_t = \frac{y_t}{y_{\text{base}}} \times 100.$$

¹⁸Since index numbers are usually quoted to the first decimal place, e.g. 107.8, we can get rid of the decimals by using a multiplying factor of 1000 instead of 100 so that the index becomes 1078 and $I_{1988} = 1000$. This is done with price indexes in New Zealand.

¹⁹More accurately, the ratio is $699/632=110.6$ which is slightly different because of rounding error in the previous answer.

We emphasize that it is important to select an appropriate base period. For example, the above company went through restructuring in 1990 which enabled them to sell the appliance for a lower and more competitive price in 1991. Because of this sudden change it would seem more appropriate to use 1991 instead of 1988 as the base year. This is done in column 4 of Table 14.4.1. Will changing the base year make any difference to the ratio of two index numbers? (Try it and see!)

The above index number is called a *price index*. Although it may be useful to the company concerned it does not contribute very much to the overall economic picture. If, however, the price was an average price over the whole country, then the index could be related to the rest of the country's economy. Clearly the base period should then be chosen at a time of relative stability in the prices when there are no major disrupting influences.

Although an index is frequently used with prices it is also used for other variables e.g. those associated with demography.

An index which is based on a single time series variable is called a *simple index number*. However in economics and business we frequently have several time series variables (in fact hundreds for an overall price index) and we wish to combine them in some way to form a composite index number. The simplest of such index numbers is obtained by simply summing all the variables to get a new variable and then constructing the index number from the new variable; the so-called *simple composite index number*. We now give an example.

Table 14.4.2 : Annual Retail Prices for Selected Items^a

Item	Price		
	1992	1993	1994
paint (per 4 litres)	67.23	60.72	58.72
hand drill	177.20	173.73	230.36
shoes (per pair)	68.32	70.60	77.67
tissues (per 200)	2.58	2.65	2.51
Sum	315.33	307.70	369.26

^aSource: Statistics New Zealand, Retail Prices of Selected CPI Items.

In Table 14.4.2 we have, for illustrative purposes, just four variables. These are the national average prices of four items, namely white acrylic house paint (per 4 liters), an electric hand drill (7 volt cordless, 10mm variable speed), medium-size school shoes (per pair) and double-thickness facial tissues (per box of 200) determined for each year from 1992 to 1994 (in February). If 1992 is the base year then the simple composite index number for 1994 is the sum of the four numbers for 1994 divided by the sum for 1992, namely

$$\frac{369.26}{315.33} \times 100 = 117.1.$$

We might ask, is it a good idea to just add the values of the variables? First of all if we change the units, for example give the cost of paint per liter (or per US gallon) and the cost of tissues per box of 150, then we would change the value of the index number for 1994. This is clearly unsatisfactory as we would expect any self-respecting index number to be independent of the units chosen. Also the cost of the hand drill is higher than the other costs and will tend to swamp any changes in the other three variables. It would be like dropping a large stone and a small pebble into the water and watch the waves from the stone swamp the ripple from the pebble. What we are saying is that large variables contribute more weight to this composite index than the other variables. This may not be too unreasonable if the large variable is the most important. However, as in this example, this is not usually the case. We would buy tissues frequently, school shoes less frequently (though probably too frequently for most parents!), some paint occasionally, and hand drills rarely (if at all). In terms of consumption, the largest variable is the least important and the smallest variable the most important. We need to attach a “loading” to each variable where the size of the loading reflects the relative importance of that particular variable. How do we choose the loadings?²⁰

One objective method is to simply use the quantity of the item consumed per person for the loading. This amounts to multiplying the cost times the consumption per person i.e. using the total amount of money per person spent on that item for a particular year. If, however, we are going to compare this with the base year it might be more appropriate to use the quantity of the item consumed per person during the base year rather than the present year. This latter approach leads to the so-called **Laspeyres price index**, namely

$$I_t(\text{Laspeyres}) = \frac{\sum_{items} (\text{quantity})_{base} (\text{price})_t}{\sum_{items} (\text{quantity})_{base} (\text{price})_{base}}$$

One advantage of this index is that the quantities of items are held constant from the base year so that we don’t need to estimate $(quantity)_t$ for each year t : all we need is $(quantity)_{base}$.

In order to apply the above formula to the data in Table 14.4.2 we need to know the consumption per person of each item in 1992. How do we estimate these? Not so easily! For the moment let us guess the 1992 consumptions per person: 1 liter of paint, one hand drill per 20 people every 5 years (i.e. $1/(20 \times 5) = 0.01$ drills per person per year), 2 pairs of shoes per 10 people (i.e. 0.2 pairs per person per year) and 13 boxes of tissues. (What are your guesses?).

²⁰We use the term “loading” instead of “weight” as we use “weight” with a different meaning later.

22 Time Series

The loadings (quantities) for the four items are then 0.25, 0.01, 0.2 and 13 respectively giving the index for 1994 as

$$\begin{aligned} I_{1994} &= \frac{0.25 \times 58.72 + 0.01 \times 230.36 + 0.2 \times 77.67 + 13 \times 2.51}{0.25 \times 67.23 + 0.01 \times 177.20 + 0.2 \times 68.32 + 13 \times 2.58} \times 100 \\ &= \frac{14.68 + 2.30 + 15.53 + 32.63}{16.81 + 1.77 + 13.66 + 33.54} \times 100 \\ &= \frac{65.15}{65.78} \times 100 = 99. \end{aligned}$$

This contrasts strongly with the unweighted index of 117.1 calculated previously. The drop is caused by the tissues now having the greatest loading and the greatest overall cost. We note that the denominator of 65.78 in the above calculations is the total amount spent on the four items per person in 1992 and the numerator of 65.78 is the same total for 1994 but based on the 1992 consumptions rather than the 1994 consumptions.

Unfortunately we still have problems with the way we calculated the weighted index. Firstly, if we try to calculate a so-called **consumer price index** based on a very large number of items (called the **market basket**) instead of just four as above we have a massive estimation problem in trying to estimate the quantities of each item consumed. Secondly, people tend to operate as households rather than individuals. Clearly one answer to both problems would be to carry out a detailed household survey across the country during the base year. Fortunately, as we now shall show, we don't need to know all the details about the quantities of each item bought by each household, only the amount spent on each item. To help us in our thinking, let us calculate the 1994 Laspeyres index for just the first two items in Table 14.4.2 using the loadings of 0.25 and 0.01 introduced above, namely

$$\begin{aligned} I_{1994} &= \frac{0.25 \times 58.72 + 0.01 \times 230.6}{0.25 \times 67.23 + 0.01 \times 177.20} \times 100 \\ &= \frac{0.25 \times 67.23 \left(\frac{58.72}{67.23} \times 100\right) + 0.01 \times 177.20 \left(\frac{230.36}{177.20} \times 100\right)}{0.25 \times 67.23 + 0.01 \times 177.20} \\ &= \frac{16.81(87.3) + 1.77(130.0)}{16.81 + 1.77} \\ &= \frac{16.81}{18.58}(87.3) + \frac{1.77}{18.58}(130.0) \\ &= 0.905(87.3) + 0.095(130.0) \\ &= 91.4. \end{aligned}$$

This seems a very roundabout way of carrying out the calculation! Let's look at what we have done. We note first of all that \$16.81 is the amount spent on

item 1 using the quantity (0.25) consumed in the base year, and this represents a proportion 0.905 of the total expenditure (\$18.58) for the two items that is spent on item 1 in the base year. Similarly \$1.77 is the amount spent on item 2 representing a proportion 0.095 of the total expenditure. Also the numbers 87.3 and 130.0 are price indexes for the two items relative to the base year. What we have shown by the above calculations is that

$$I_{1994} = w_1 I_{1994}^{(1)} + w_2 I_{1994}^{(2)},$$

where $w_1 + w_2 = 1$, w_i is the proportion the total expenditure spent on item i in the base year and $I_{1994}^{(i)}$ is the price index for item i ($i = 1, 2$).²¹

If we used all four items in Table 14.4.2 we would have

$$I_{1994} = w_1 I_{1994}^{(1)} + w_2 I_{1994}^{(2)} + w_3 I_{1994}^{(3)} + w_4 I_{1994}^{(4)},$$

where $w_1 + w_2 + w_3 + w_4 = 1$. In words, the index I_t can be found by calculating a price index for each item and then taking a weighted sum of these indexes where the weight attached to an index is the proportion of total expenditure that is spent on that item in the base year. The weights add to 1.

What does the above alternative formulation mean in practice when it comes to calculating I_t ? Firstly, for each item, w_i and $I_t^{(i)}$ can be estimated from different sources. For example w_i could be found from a household expenditure survey in which a random sample of households is asked to keep track of how much they spend on various items over a sample period in the base year: the actual quantity of each item bought does not need to be recorded. On the other hand, the price index $I_t^{(i)}$ for the item can be obtained from a separate survey of prices for the item and taking a suitable average: we don't need to know the quantities of the various items consumed. Secondly, items can be grouped together, e.g. fruit, provided we can calculate a price index for that group. For example, in Table 14.4.3 we have some average expenditure data per household.²² We will construct a consumer price index for food based on the expenditure and individual price index information given in the table.

In Table 14.4.3 the individual indexes are with respect to a base year of 1988. To get an overall food price index using 1992 as the new base year, we get I_{1994} by multiplying corresponding elements in column (2) by the elements obtained from dividing 100 times column (4) by column (3), and then adding. We thus

²¹Mathematically, if the subscript zero represents the base year, and q and p represent quantity and price respectively, we can express I_t in the form

$$I_t = \frac{\sum q_0 p_t}{\sum q_0 p_0} = \frac{\sum q_0 p_0 \frac{p_t}{p_0}}{\text{Total cost}} = \sum \frac{\text{Item cost } p_t}{\text{Total cost } p_0} = \sum \frac{\text{Item cost}}{\text{Total cost}} I_t^{(Item)} = \sum_i w_i I_t^{(i)}.$$

²²Here Fruit & Veg. includes fresh, frozen, dried and canned goods; Grocery Foods includes soft drinks and confectionery; and Meals Out includes restaurant meals, takeaways and ready-to-eat food.

24 Time Series

get

$$\begin{aligned}
 I_{1994} &= 0.1555 \left(\frac{106.2}{106.9} \times 100 \right) + 0.1844 \left(\frac{98.2}{93.9} \times 100 \right) + 0.4696 \left(\frac{99.7}{98.3} \times 100 \right) \\
 &\quad + 0.1904 \left(\frac{100.2}{98.6} \times 100 \right) \\
 &= 0.1555(99.35) + 0.1844(104.58) + 0.4696(101.42) + 0.1904(101.62) \\
 &= 101.7.
 \end{aligned}$$

What about I_{1992} ? Using 1992 as the base year means that we effectively use $I_{1992}^{(i)} = 100$ for each group so that I_{1992} will be $100 \times$ (sum of weights) i.e. $I_{1992} = 100$ also. Thus $I_{1994} = 101.7$ means that the cost of food has risen by 1.7% from 1992 to 1994.

Table 14.4.3: Household Data for Constructing Food Price Indexes^a

Food Group	March 1992		Price Index (3)	March 1994
	Average weekly Expenditure (\$) (1)	Weight (w) (2)		Price Index (4)
Fruit & Veg.	15.60	0.1555	106.9	106.2
Meat, Fish & Poultry	18.50	0.1844	93.9	98.2
Grocery Foods	47.10	0.4696	98.3	99.7
Meals Out	19.10	0.1904	98.6	100.2
Sum	100.30	1.000		

^aSource: Statistics New Zealand.

One can develop similar indexes for other groups of items. In most countries the items are usually divided up into groups such as food, housing, household operation (power, furnishings etc.), apparel, transport, recreation etc. These are further subdivided into subgroups as we have done in Table 14.4.3. Weights for each of the groups are found from household and other surveys and the separate group indexes are then combined as in the construction of the food index to get a combined index for all groups called the Consumers Price Index (CPI). We see that that the CPI is found by a repeated process of weighting and adding with respect to items, subgroups and groups. In addition to keeping track of the CPI, indexes for some of the groups and subgroups are followed to see what changes are occurring in various sectors of the economy.²³

²³In New Zealand separate indexes are calculated for 8 commodity groups and 21 commodity subgroups. As previously mentioned, a multiplying factor of 1000 is used instead of 100 to remove decimal points.

The above description is obviously an oversimplification. For example what do we do about price variations and different brands, regional variations, and variations in the size of a household? These problems can also be dealt with using suitable weighting methods. We won't go into details. Also there are seasonal changes in the pattern of household spending (e.g. power has a greater weight in winter) so that the CPI is computed quarterly.

Unfortunately consumer spending patterns change²⁴ so that the consumption weightings will change with time. We could allow for this by replacing $(quantity)_{\text{base}}$ by $(quantity)_t$ everywhere in the Laspeyres index to get another index called the Paasche price index. This amounts to using weights based on information from year t rather than from the base year. The weights then have to be recalculated every year and this would involve a lot of surveying. For this reason we shall stay with the Laspeyres index as it is much easier to calculate and, therefore, more widely used. We can take care of changes in the pattern of consumer spending by changing the base year from time to time and carrying out extensive surveys in the base year to get the correct weightings.²⁵

It should be emphasized that the above indexes are essentially price expenditure indexes. Other price indexes however are also produced.²⁶

How can the CPI index be used? With most nations it is used as (i) a general measure of inflation, (ii) an indicator of how price changes effect private households, (iii) a deflator to produce constant price equivalents for time series featuring or representing actual expenditure values (e.g. for converting the prices of a particular item the years 1990 to 1994 into 1990 equivalents to see if the price had changed relative to general inflation during this period), and (iv) as an indicator to see if benefits, allowances etc. are keeping up with the general inflation. Although the CPI does not strictly fit any of the above applications it is a useful measure of the cost of living. For example, if the CPI is 104.2 in 1994 and the base period is 1992, then the CPI has gone up 4.2% since 1992. It is usually stated that the cost of living has gone up 4.2%, from 1992 to 1994. By the same token, the purchasing power of the dollar in 1994 compared with that in 1992 is $\frac{100}{104.2}$ dollars or \$0.96.

Quiz for Section 14.4

1. What is the difference between a single index number and a simple composite index number?
2. Why is it inappropriate to just add the values of the variables when combining them in a composite index number?

²⁴due to such factors as changes in relative prices and the availability of goods and services, changes in fashions and taste, changes in the size and distribution of the population, and changes in retailing patterns.

²⁵In New Zealand some previous base years have been 1983, 1988 and 1993; all in December.

²⁶In New Zealand we have the Producers Price Index, the Capital Goods Price index, the prevailing Weekly Wage Rates Index and customized cost-adjustment indexes for specific industries, companies and government departments.

3. What important advantage does the Laspeyres index number have?

Exercises on Section 14.4

1. Using the data in Table 14.4.1 on the annual prices of an appliance, obtain a price index series using 1990 as the base year. Using just your series, find what percentage the price has gone up (a) from 1990 to 1993, (b) from 1991 to 1993.
2. It is common practice to compute a combined price index for tobacco and alcohol. In the table below we have the average household expenditure on these two items for March 1992, along with their price indexes for the March quarter (using a multiplying factor of 1000 instead of 100 to remove decimals). Construct a Laspeyres index combining the two items for 1992, 1993 and 1994.

Items	March 1992 Average weekly Expenditure (\$)	Price Index		
		1992	1993	1994
Tobacco products	8.50	974	992	999
Alcoholic drinks	14.70	976	990	1002

Source: Statistics New Zealand.

14.5 Summary

1. A time series is best graphed by joining up consecutive points in time.
2. A time series can exhibit positive or negative **correlation** depending on whether consecutive pairs of points tend to follow each other or tend to move in opposite directions. Unless there is some compensatory mechanism taking place, the correlation is generally positive.
3. Monthly or quarterly variation can be removed by using annual totals.
4. Quarterly data can either be plotted with the first quarter plotted at the beginning of the year or the fourth quarter plotted at the beginning of the next year.
5. A time series may contain a trend, one or more cycles and irregular variation.
6. We need to watch out for *spikes* and *steps*.
7. A **3-point moving average** is obtained by averaging the observations y_t three at a time i.e. by averaging observations one, two and three, then observations two, three and four, and so on. There are no smoothed points for the first and last time periods.

8. A ***k*-point moving average** is obtained by averaging the observations k at a time. When k is odd (say $k = 2m - 1$, where m is a positive integer), the smoothed series is in phase with the original series i.e. its time periods are the same as the original series. There are no smoothed points for the first and last m points. When k is even (say $k = 2m$), the smoothed series has time periods midway between those of the original series.
9. **Quarterly variation** can be smoothed using a 4-point moving average. The smoothed series can be brought back into phase with the original series, i.e. **centered**, by using a further 2-point moving average.
10. Any cycle covering c time periods (with period c) can be smoothed using a c -point moving average.
11. A linear trend can be removed by “differencing” the original series to get $z_t = y_{t+1} - y_t$.
12. A time series can be **deseasonalized** as follows: (i) Average over all the y_t for each quarter to get 4 numbers s_1, s_2, s_3 and s_4 . (ii) Subtract off the mean \bar{s} of these 4 numbers from each of them to get $s_t - \bar{s}$. (iii) Calculate $y_t - (s_t - \bar{s})$, where the definition of s_t is extended beyond the first year by simply repeating the same 4 numbers. It is assumed that the underlying model is additive for y_t , otherwise use $\log y_t$ for a multiplicative model.
13. An index number measures the change in a time series variable y_t relative to a **base year**. Mathematically it is given by

$$I_t = \frac{y_t}{y_{\text{base}}} \times 100.$$

14. The Laspeyres index number is a composite index number defined by

$$I_t(\text{Laspeyres}) = \frac{\sum_{\text{items}} (\text{quantity})_{\text{base}} (\text{price})_t}{\sum_{\text{items}} (\text{quantity})_{\text{base}} (\text{price})_{\text{base}}}.$$

15. The **consumer price index** is a Laspeyres index number based on a large number of items called the **market basket**. It is calculated by multiplying the proportion of the weekly budget spent on a particular item by the price index for that item and then summing the answer over all the items in the basket.

Review Exercises 14

- Table 1 gives the marriage rates and birth rates (per 1,000 of population) for each of 3 countries from 1974 to 1993.

Table 1 : Marriage Rates and Birth Rates^a

Year	Marriage Rates ^b			Birth Rates ^c		
	NZ	UK	US	NZ	UK	US
1974	8.4	15.6	10.5	19.5	13.2	14.8
1975	7.9	15.4	10.0	18.3	12.5	14.6
1976	7.7	14.5	9.9	17.6	12.1	14.6
1977	7.2	14.5	9.9	17.2	11.8	15.1
1978	7.1	14.9	10.3	16.2	12.3	15.0
1979	7.1	14.9	10.4	16.7	13.1	15.6
1980	7.3	15.0	10.6	16.1	13.5	15.9
1981	7.5	14.1	10.6	16.1	13.0	15.8
1982	8.0	13.4	10.6	15.7	12.8	15.9
1983	7.7	13.8	10.5	15.7	12.8	15.6
1984	7.8	14.0	10.5	15.9	12.9	15.6
1985	7.5	13.9	10.1	15.8	13.3	15.8
1986	7.3	13.9	10.0	16.1	13.3	15.6
1987	7.4	14.0	9.9	16.7	13.6	15.1
1988	7.1	13.8	9.8	17.4	13.8	16.0
1989	6.8	13.7	9.7	17.4	13.6	16.4
1990	6.9	13.1	9.8	17.9	13.9	16.7
1991	6.8	12.1	9.4	17.6	13.7	16.3
1992	6.4	12.3	9.3	17.2	13.5	15.9
1993	6.3	11.7	9.0	16.9	13.1	15.5

^aSources: Statistics NZ, US Dept of Commerce, UK Central Statistical Assoc., UK Annual Abstract of Statistics.

^bMarriages per 1,000 of pop. ^cBirths per 1,000 of pop.

- What factors do you think affect the birth rate? the marriage rate?
- Would you expect any relationship between the birth rate and marriage rate? If so what? What social trends may have weakened any such relationship?
- Plot the NZ birth-rate time series. Comment on what you see.
- If you are using a computer, plot all 3 marriage-rate series on the same graph. Comment on any trends, patterns, similarities, differences etc. Working from the plots by eye, predict the 1994 marriage rates for each of the 3 countries.
- Repeat (d) for the birth-rate series.
- There are quite large differences in marriage rates between countries. What are some possible causes of this? Are the birth-rate patterns relevant to any of your ideas?
- Using the US marriage-rate data from 1984 to 1993, and a base year of 1989, obtain an index for the marriage rates. How much has the marriage rate gone down (i) from 1989 to 1993, (ii) from 1991 to 1993?

2. Table 2 gives CPI figures for 5 countries from 1985 to 1994. Each index is based on the December Quarter 1993 (1000).

Table 2 : CPI for 5 Countries^a

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
Australia	606	654	714	770	826	891	949	972	981	999
Canada	710	739	769	803	836	880	924	964	980	995
NZ	542	625	716	811	853	908	958	974	984	998
UK	637	675	696	724	768	828	908	951	981	998
US	719	744	756	787	821	861	908	941	970	997

^aSource: Statistics New Zealand, 1994.

- (a) Plot the time series for the UK. Comment on any features.
- (b) The CPI for the UK in 1984 was 606. For each year, compute the % increase over the previous year. This is a measure of inflation for that year. Plot this new time series. Compare it with Fig. 14.2.1 for the US.
- (c) Smooth the plot in (b) using a 3-point moving average. What do you conclude?
- (d) Repeat (a)–(c) for any other country you are interested in. The 1984 index figures were 585 (Aust.), 683 (Can.), 499 (NZ) and 691 (US).
- (e) If you are using a computer, plot all 5 time series on a single graph. What do you learn from this plot?
3. This exercise uses the global warming data of Table 14.1.2.
- (a) Plot the series and summarize any trend with a freehand trend curve (drawn by eye).
- (b) (Computer only, using Chapter 12 ideas.) Suppose that we wanted to summarize the trend we see. Does a quadratic curve do this job adequately? Try it and see. Superimpose the least squares quadratic curve on the plot from (a). Plot residuals versus year.
- (c) Exercises 14.2.1, problem 3, described differencing as a way of removing linear trend. How well does this technique work here? Plot the time series of the differences $Z_t = y_{t+1} - y_t$ and comment.
4. Table 3 gives quarterly unemployment figures for the period 1993–1996 for each of 5 countries. Choose a country of interest and perform the following.
- (a) Plot the unemployment series for the country you have chosen. Superimpose the 4-point moving average. How well does it smooth the data and highlight the trend?
- (b) Estimate the seasonal adjustment terms $s_i - \bar{s}$ for each of the four quarters. What do these terms tell you?
- (c) Deasonalize the data and plot the resulting series. Does this tell you anything new?

- (d) If you are using a computer, plot the deseasonalized series for each of the 5 countries on the same graph. Obtain the seasonal adjustments for each of the 5 countries. What does this information tell you about comparative unemployment patterns?

Table 3: Unemployment Rates for 5 Countries^a

	1993				1994				1995				1996			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Aus.	11.8	10.7	10.6	10.5	11.2	9.8	9.2	8.8	9.6	8.3	8.1	8.2	9.2	8.3	8.4	8.4
Can.	12.0	11.4	10.9	10.7	12.0	10.7	9.8	9.2	10.6	9.5	9.2	8.9	10.4	9.6	9.4	9.4
NZ	10.3	9.9	9.5	9.5	9.6	8.4	7.7	7.5	6.9	6.2	5.9	6.2	6.5	5.9	6.1	6.0
UK	10.7	10.4	10.4	9.9	10.1	9.5	9.4	8.7	8.7	8.3	8.3	7.9	8.2	7.7	7.7	6.8
US	7.6	6.9	6.6	6.2	7.1	6.1	5.9	5.3	5.9	5.6	5.6	5.2	6.0	5.4	5.2	5.0

^aSources: Several issues of *Main Economic Indicators*, Statistics Directorate of the OECD.

5. An icecream manufacturer's monthly sales in thousands of dollars were recorded every month from April 1993 to March 1997. The data is given in Table 4.

Table 4: Monthly Icecream Sales (\$000's)

Month	1993	1994	1995	1996	1997
Jan	-	470	548	589	556
Feb	-	443	506	561	658
Mar	-	386	444	493	557
Apr	327	342	448	461	-
May	288	319	408	430	-
Jun	269	307	387	434	-
Jul	256	284	373	406	-
Aug	286	309	408	428	-
Sep	298	326	420	450	-
Oct	318	359	429	481	-
Nov	329	376	442	485	-
Dec	381	416	499	536	-

- (a) Plot the time series. Comment on any features. Summarize any trend with a freehand curve. If you see cycles, comment upon the positions of peaks and troughs and whether the amplitude (the distance from the bottom of a trough to the top of a peak) is changing or seems to be fairly constant.
- (b) Now smooth your time series using a 4-point moving average. Is your new series helpful?
- (c) Smooth the original plot using a 12-point moving average. What do you conclude? What sales would you predict for the next 12 months?
- (d) Form a new quarterly series by adding together the months January, February and March; then April, May and June; July, August and September; and finally October, November and December. Plot this series and then smooth it using centered 4-point smoothing. Compare this series with the smoothed series in (b).

- (e) Deseasonalize the time series that you obtained by adding the months together in (d). Have you learnt anything new?
6. A small business in Auckland sells pamphlets. The monthly income in dollars from January 1994 to May 1997 is given in Table 5.

Table 5 : Monthly Pamphlet Sales (\$)

Month	1994	1995	1996	1997
Jan	2045	2967	3579	3783
Feb	3545	3650	3328	4785
Mar	4607	4131	5087	5141
Apr	2143	3896	5144	7975
May	3909	4951	4584	7019
Jun	3919	4396	4940	-
Jul	3134	5142	4130	-
Aug	2934	3566	4672	-
Sep	3294	4698	5333	-
Oct	3651	3814	4443	-
Nov	5101	7085	9826	-
Dec	4792	6565	6598	-

- (a) Plot the time series. Comment on any special features of the series such as variability, special points, and the presence or absence of any seasonal effect. If you see cycles comment upon the positions of peaks and troughs and whether the amplitude (the distance from the bottom of a trough to the top of a peak) is changing or seems to be fairly constant.
- (b) Smooth the series using a three-point moving average. Is this new series helpful?
- (c) What other smoothing would you recommend? Try your ideas out and see how well they work?
7. The possibility of cyclic behavior in animal populations has always been of interest to ecologists. Table 6 contains a very famous set of data giving the annual number of lynx pelts, and the cost per pelt, for pelts bought by Hudson's Bay Company in Canada from 1857 – 1911 inclusive.
- (a) What factors, apart from lynx population-numbers, might affect the numbers trapped?
- (b) Plot the time series for the number of pelts. Do you think there is a population cycle? If so, where are the peaks?
- What affect do the numbers of pelts have on the price?
- (c) Plot the time series for the cost per pelt. Compare this with the times series plotted in (b). What do you think is happening?
- (d) Plot the number of pelts versus cost per pelt for each year up until 1900. Does this shed any light on a relationship between numbers and cost?
- (e) In (d), we did not use the last part of the series. What, if anything, does the behavior of the last part of the cost series suggest to you?

Table 6 : Numbers of Lynx Pelts^a

Year	No. ^b	Cost ^c	Year	No.	Cost	Year	No.	Cost	Year	No.	Cost
1857	23362	12.75	1871	15686	6.50	1885	27187	11.67	1899	26761	10.08
1858	31642	8.08	1872	7942	11.87	1886	51511	18.75	1900	15185	27.58
1859	33757	9.50	1873	5123	19.87	1887	74050	9.75	1901	4473	17.25
1860	23226	10.00	1874	7106	14.33	1888	78773	9.33	1902	5781	29.50
1861	15178	8.08	1875	11250	14.17	1889	33899	19.08	1903	9117	45.67
1862	7272	8.50	1876	18774	13.00	1890	18886	13.75	1904	19267	24.33
1863	4448	12.25	1877	30508	8.25	1891	11520	13.58	1905	36116	27.00
1864	4926	14.50	1878	42834	7.42	1892	8352	19.33	1906	58850	26.83
1865	5437	12.58	1879	27345	8.58	1893	8660	16.58	1907	61478	27.25
1866	16498	11.58	1880	17834	10.08	1894	12902	11.50	1908	36300	38.43
1867	35971	8.17	1881	15386	12.58	1895	20331	12.08	1909	9704	87.33
1868	76556	6.83	1882	9443	14.67	1896	36853	7.75	1910	3410	123.08
1869	68392	6.09	1883	7599	16.75	1897	56407	6.17	1911	3774	102.33
1870	37447	5.50	1884	8061	19.17	1898	39437	7.08			

^aSource: Andrews and Herzberg [1985].^bNumber of lynx pelts bought. ^cCost per pelt (\$).

8. Table 7 gives measures of annual sunspot activity for the 40 years 1916–1955 inclusive.

Table 7 : Sunspot Activity, 1916–1955^a

Year	Activ.								
1916	57.1	1924	16.7	1932	11.1	1940	67.8	1948	136.3
1917	103.9	1925	44.3	1933	5.7	1941	47.5	1949	134.7
1918	80.6	1926	63.9	1934	8.7	1942	30.6	1950	83.9
1919	63.6	1927	69.0	1935	36.1	1943	16.3	1951	69.4
1920	37.6	1928	77.8	1936	79.7	1944	9.6	1952	31.5
1921	26.1	1929	64.9	1937	114.4	1945	33.2	1953	13.9
1922	14.2	1930	35.7	1938	109.6	1946	92.6	1954	4.4
1923	5.8	1931	21.2	1939	88.8	1947	151.6	1955	38.0

^aSource: Rao and Gabr [1984].

- (a) Plot the time series. What do you see? Summarize any trend using a freehand curve. If you see cycles, comment upon the locations of the peaks and the behavior of the amplitudes of the cycles?
- (b) Smooth the time series using a 10–point moving average. What does this confirm?
9. It was conjectured that increases in labor costs resulted in increases in the market share of dry wall construction in the construction market. Table 8 gives a measure of this market share, called “market penetration”, for 3 countries. Estimated hourly labor costs and annual inflation figures are also given for the years 1974–1991.
- (a) Separately, plot market penetration versus labor costs for each country. Does this appear to support the conjecture?

There was a great deal inflation over this time period. We shall investigate the relationship between market penetration and labor costs when

labor costs are measured in constant (1974) currency values (yen or dollars depending upon country).

- (b) For each country, set up a currency index with a base of 1 in 1974 to represent the inflated value of a unit of 1974 currency. For example, by the beginning of 1975, a 1974 NZ dollar has inflated by 11.1% to be worth \$1.111, and by another 14.6% to be worth $\$1.111 \times 1.146 = \1.273206 by the beginning of 1976, and so on.
- (c) For each country, convert labor costs to constant 1974 currency values by dividing the printed labor cost by the index value (i.e. by the inflated value of the currency unit). We will call these “indexed labor costs.”
- (d) For each country, plot market penetration versus the indexed labor costs. Does penetration still appear to increase with labor costs?
- (e) For each country, plot labor costs versus time. What do you see?
- (f) Do you think the increased market penetration is a result of increased labor costs? Support your answer.
- (g) We have indexed labor costs using the general CPI inflation figures. Can you think of a way of proceeding that might be more relevant to the building industry?

Table 8: Market Penetration of Dry-wall Construction^a

Year	Japan			New Zealand			United States		
	Penetr. ^b	Labor ^c	Inflat. ^d	Penetr.	Labor	Inflat.	Penetr.	Labor	Inflat.
1974		1386	23.1	1.91	2.25	11.1		6.81	11.0
1975		1580	11.8	2.30	2.55	14.6		7.31	9.1
1976		1776	9.4	2.21	2.89	17.0		7.71	5.7
1977	1.05	2003	8.2	2.28	3.26	14.3	4.36	8.10	6.5
1978	1.16	2188	4.1	2.42	3.64	12.0	4.01	8.66	7.6
1979	1.26	2353	3.8	2.47	4.24	13.7	4.28	9.27	11.3
1980	1.40	2516	7.8	2.46	5.01	17.1	4.56	9.94	13.5
1981	1.43	2717	4.9	2.43	6.12	15.3	4.92	10.80	10.3
1982	1.64	2796	2.7	2.46	7.17	16.2	5.35	11.63	6.2
1983	1.77	2893	1.9	2.68	7.38	7.4	4.96	11.94	3.2
1984	1.78	3041	2.2	2.87	7.65	6.2	5.05	12.13	4.3
1985	1.85	3062	2.0	2.59	8.03	15.4	5.11	12.32	3.6
1986	1.87	3197	0.6	2.57	9.86	13.2		12.48	1.9
1987	1.90	3314	0.1	2.84	10.82	15.7		12.71	3.7
1988	2.07	3484	0.7	2.81	11.74	6.4		13.08	4.0
1989	2.04	3732	2.3	3.25	12.19	5.7		13.54	4.8
1990	2.04	4016	3.1	3.49	12.85	6.1		13.78	5.4
1991	2.25	4246	3.3	3.77	13.26	2.6		14.01	4.2

^aSource: Data courtesy of Ton Hulsbosch, Fletcher Challenge Corp.

^bMarket penetration ^cLabor cost (per hour) ^dAnnual % inflation.