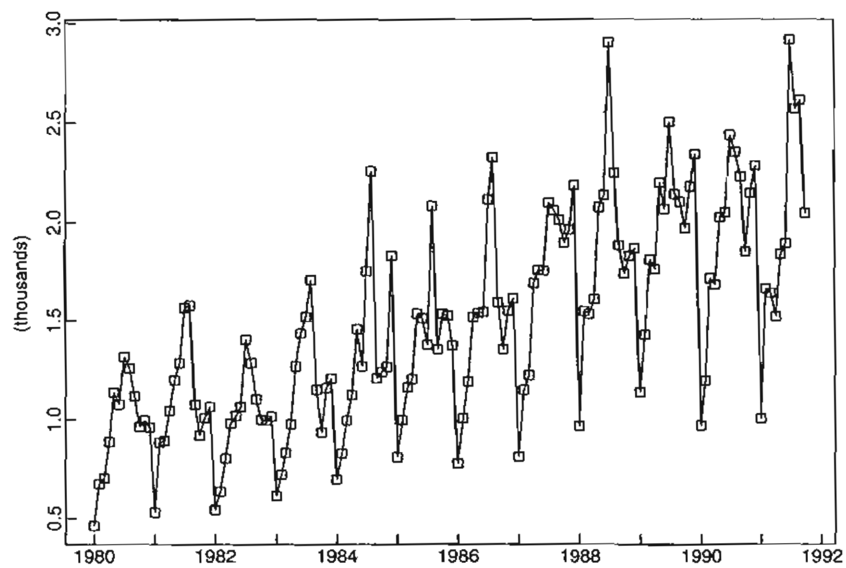

Figure 1-1
The Australian red wine sales, Jan. '80 – Oct. '91.

frequency: 1 month

$t = 1, \ldots, T$ where $T = 124$

unit of measmement: 1 kiloliter

TREND: YES, upward, linear

seasonality: YES : peak in July
trough in January

2

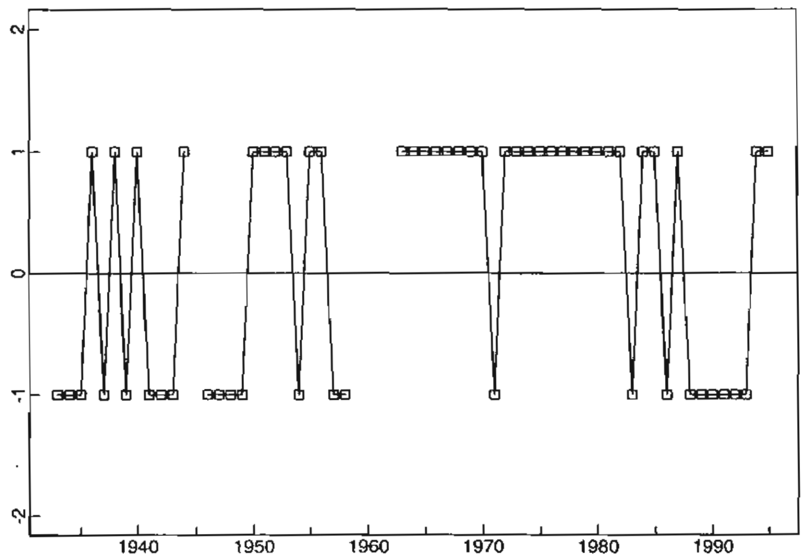

Figure 1-2
Results of the all-star baseball games, 1933–1995.

This is a series with only two possible values, ± 1. It also has some missing values, since no game was played in 1945, and two games were scheduled for each of the years 1959–1962. □

All-star baseball games, 1933–1995

Figure 1.2 shows the results of the all-star games by plotting x_t, where

$$x_t = \begin{cases} 1 & \text{if the National League won in year } t, \\ -1 & \text{if the American League won in year } t. \end{cases}$$

frequency: not constant in the sampling period! data sampled at unequal intervals

3

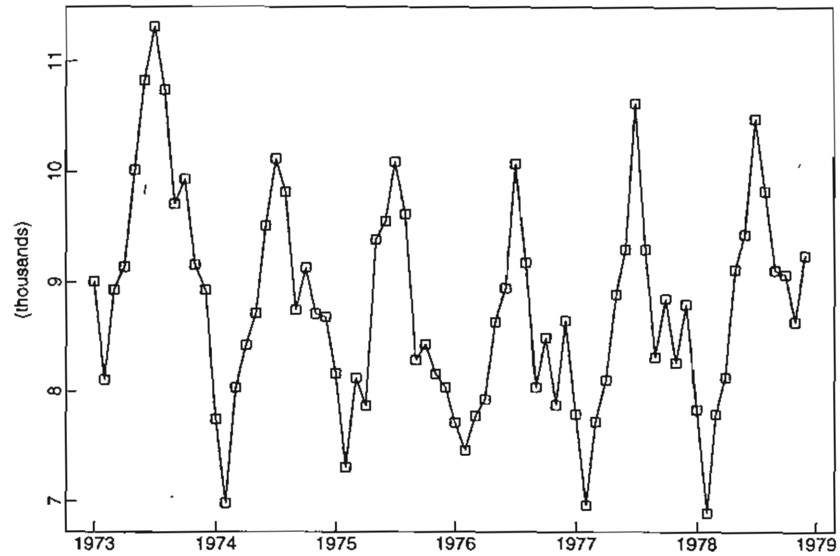

Figure 1-3
The monthly accidental deaths data, 1973–1978.

frequency : 1 month

TREND : No

SEASONALITY: YES, peak in July
trough in February

4.

frequency:
10 years

TREND : YES
nonlinear - expo
or quadratic

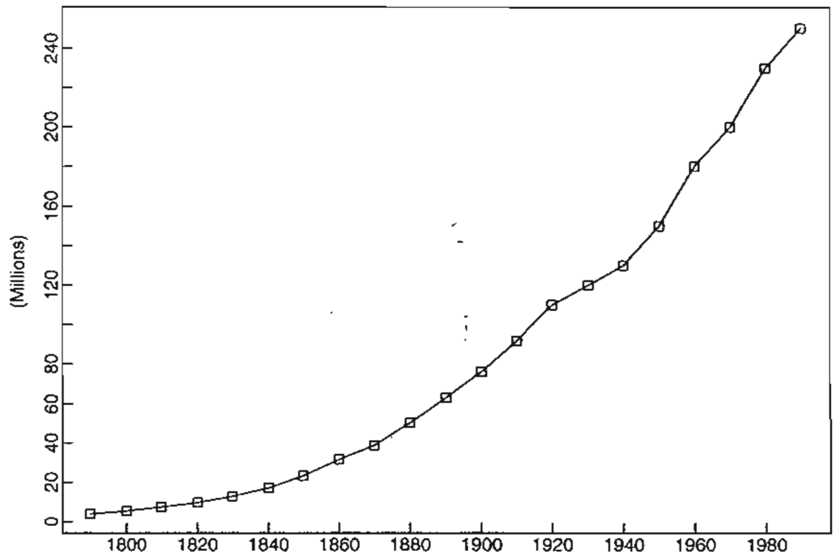

Figure 1-5
Population of the U.S.A. at ten-year intervals, 1790–1990.

frequency:
1 year

LOCAL TREND

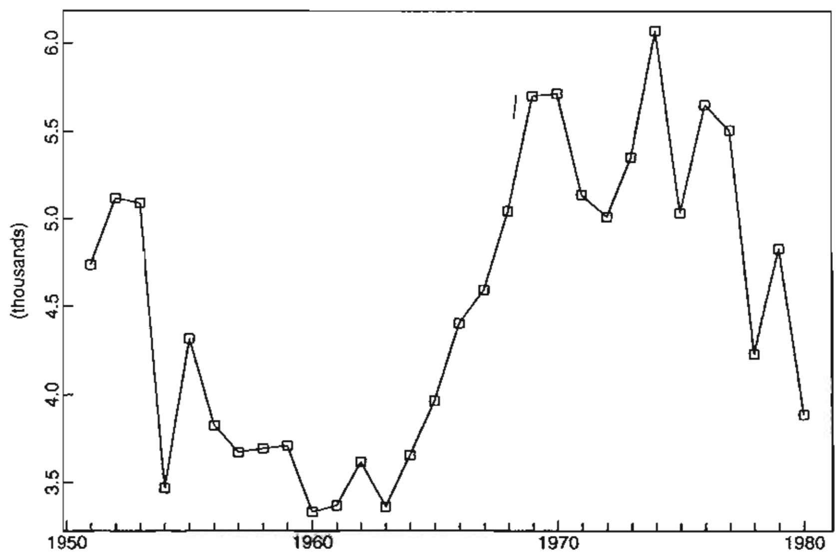

Figure 1-6
Strikes in the U.S.A., 1951–1980.

WHITE NOISE :

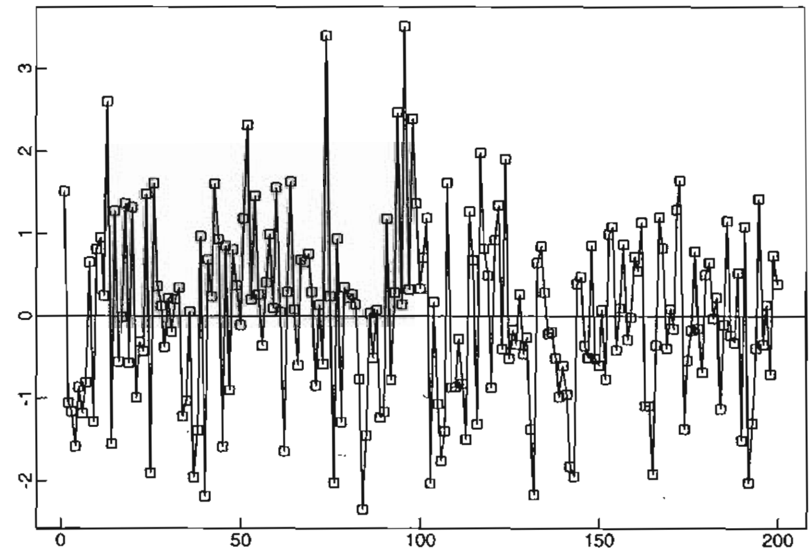

Figure 1-12
200 simulated values of iid N(0,1) noise.

TO MODEL A TIME SERIES, WE NEED TO REMOVE THE TREND AND SEASONALITY

- Suppose $$X_t = m_t + Y_t$$

m_t : trend

Y_t : a constant mean component

trend m_t can be a linear function of time

ex $$m_t = a_0 + a_1 t$$

$$X_t = a_0 + a_1 t + Y_t \quad , \quad t = 1, 2 \ldots T$$

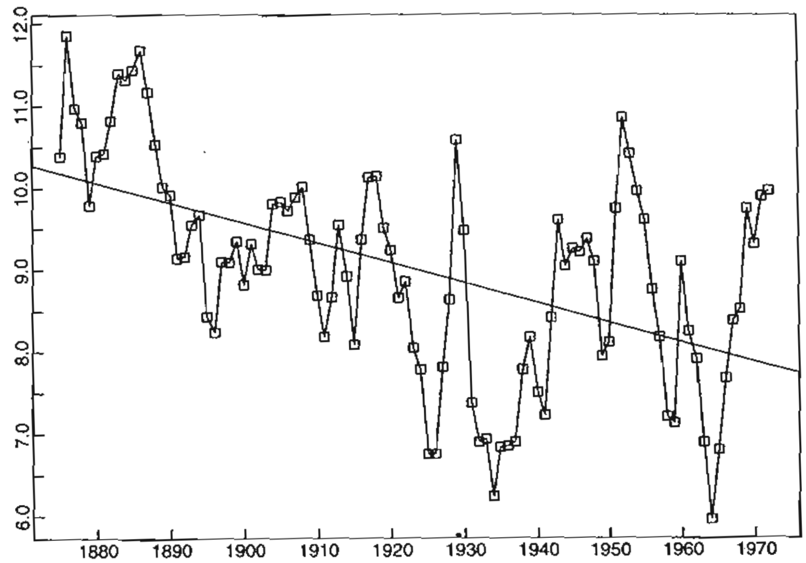

Figure 1-9
Level of Lake Huron 1875–1972 showing the line fitted by least squares.

regress by OLS X_t on a constant and t:

$$X_t = 10.202 + (-0.0242) \cdot t + \hat{Y}_t$$

the residual $\hat{Y}_t$ is a stationary component:

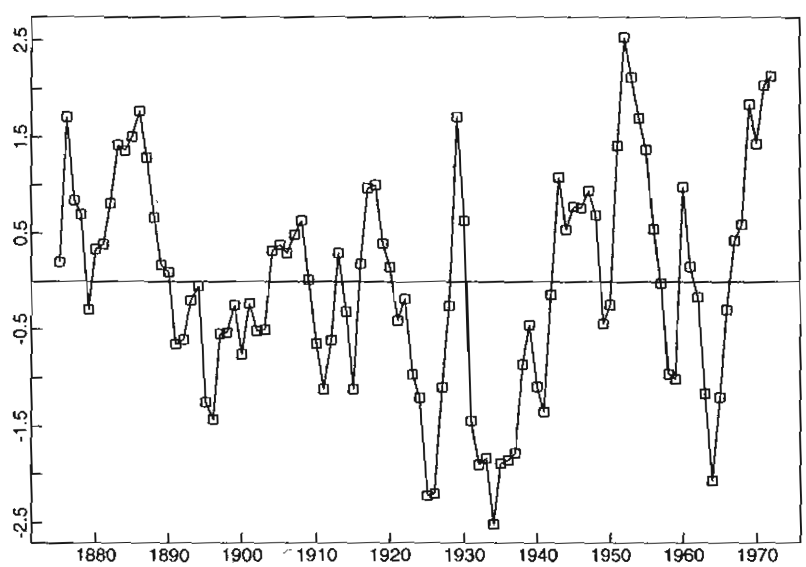

Figure 1-10
Residuals from fitting a line to the Lake Huron data in Figure 1.9.

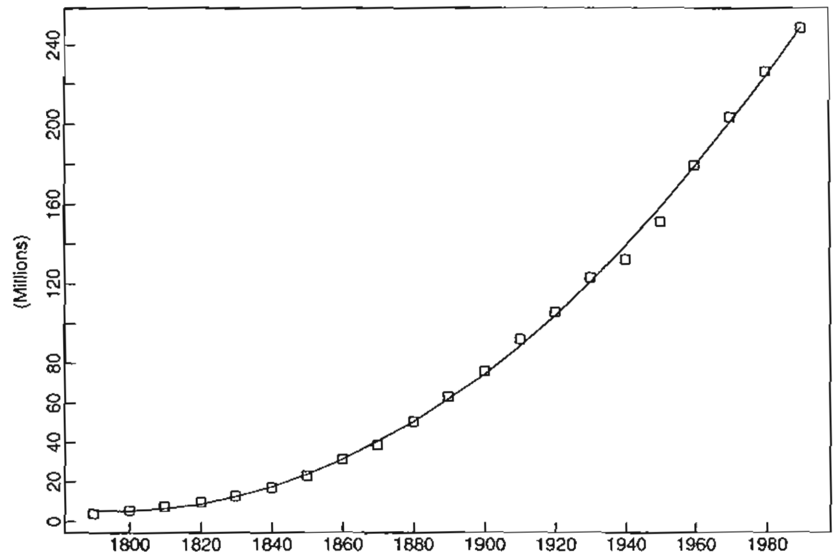

Figure 1-8
Population of the U.S.A. showing the quadratic trend fitted by least squares.

$$X_t = a_0 + a_1 t + a_2 t^2 + Y_t$$

OLS estimation $\Rightarrow$

$$\hat{X}_t = 6.95\,(10^6) + (-2.15\cdot 10^6)\,t + (6.5\cdot 10^5)\,t^2$$

- Suppose $\{X_t\}$ is a time series with finite variance.

we define

the mean $\quad \mu_t = E(X_t)$

the covariance function

$$\gamma(t+h, t) = \mathrm{cov}(X_{t+h}, X_t)$$
$$= E\left[(X_{t+h} - \mu_{t+h})(X_t - \mu_t)\right]$$

STATIONARITY

X_t IS (WEAKLY) STATIONARY IF

(a) the mean is time invariant

$$\mu_t = \mu \qquad \text{for all } t$$

(b) the covariance depends on the time between two realisations:

$$\gamma(t+h, t) = \gamma(h)$$

$\Rightarrow$ $\gamma(0)$ = variance of X_t

+ the variance is time invariant too

- Serial (temporal) dependence
- memory

autocorrelation function

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

unit free
comparable

BASIC DESCRIPTIVE STATISTICS FOR TIME SERIES:

- the mean is estimated from

$$\bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t$$

- the sample autocovariance

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

the sample variance

$$\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2$$

- the sample ACF (autocorrelation function)

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

10

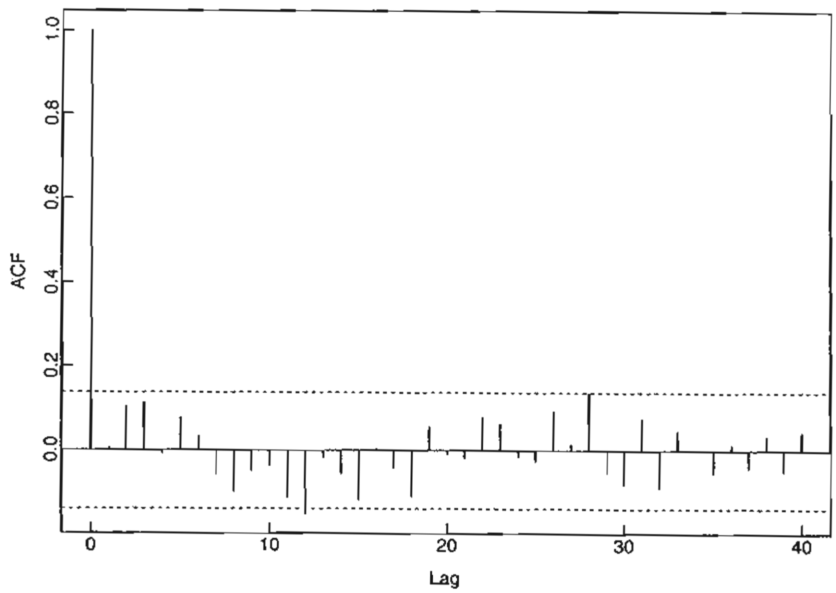

Figure 1-13
The sample autocorrelation function for the data of Figure 1.12 showing the bounds $\pm 1.96/\sqrt{n}$.

White noise

"no memory" all autocorrelations statistically = 0

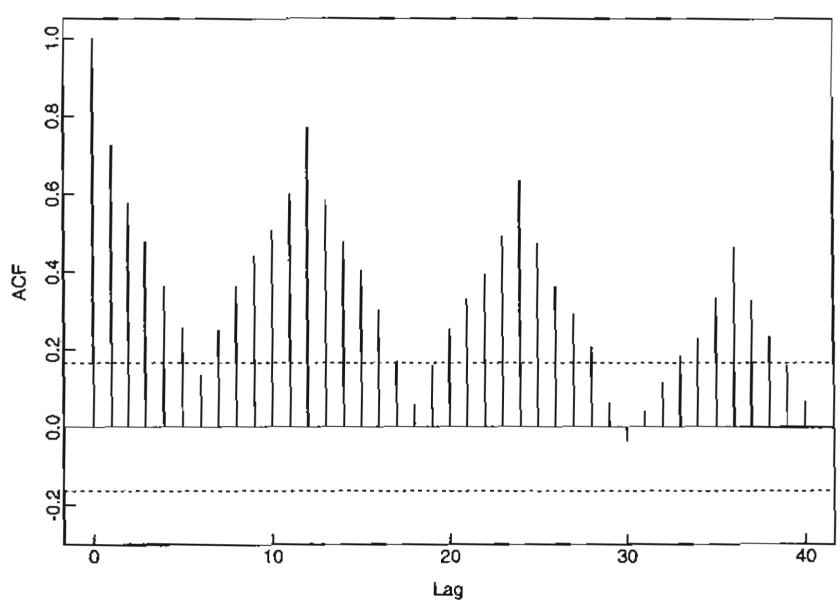

Figure 1-14
The sample autocorrelation function for the Australian red wine sales showing the bounds $\pm 1.96/\sqrt{n}$.

very(!) long memory with periodic patterns

OTHER METHODS OF TREND ELIMINATION

- moving average filter

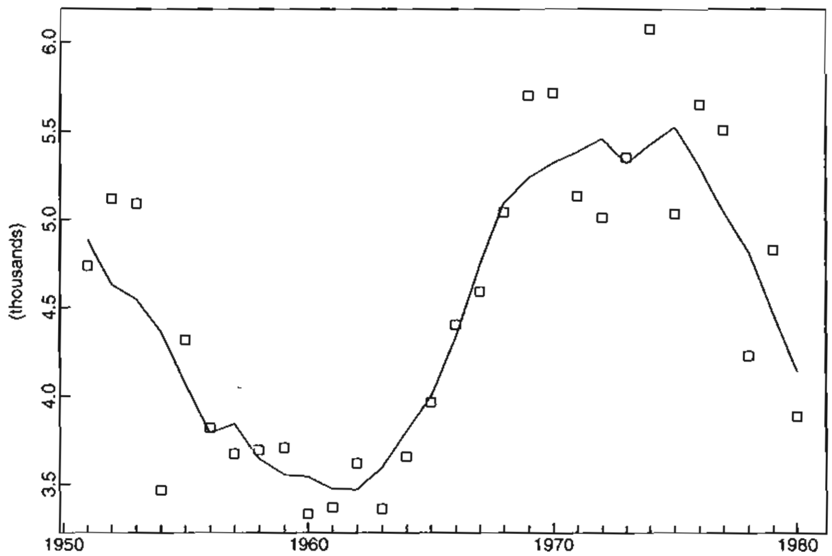

Figure 1-18
Simple 5-term moving average $\hat{m}_t$ of the strike data from Figure 1.6.

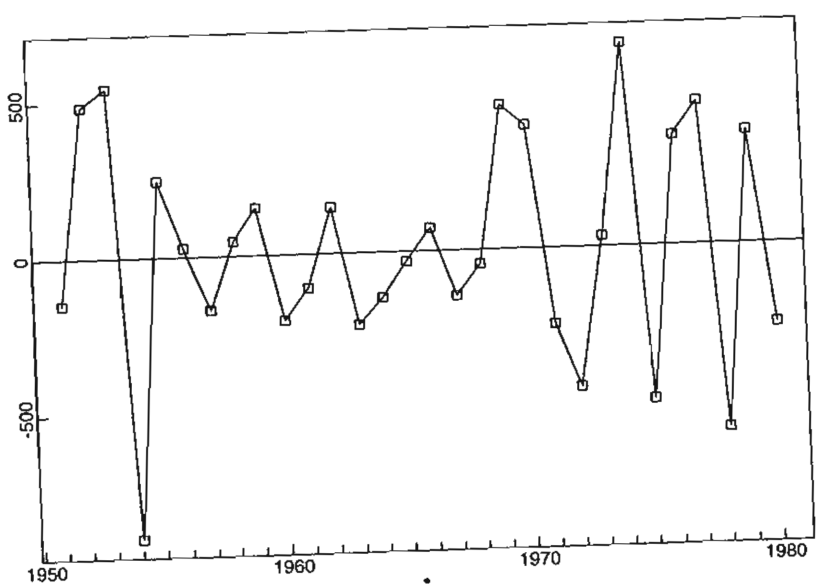

Figure 1-19
Residuals $\hat{Y}_t = X_t - \hat{m}_t$ after subtracting the 5-term moving average from the strike data

DIFFERENCING

difference operator:

$$\nabla x_t = x_t - x_{t-1}$$

$$\nabla^2 x_t = \nabla(\nabla x_t) = \nabla(x_t - x_{t-1})$$

$$= x_t - 2x_{t-1} - x_{t-2}$$

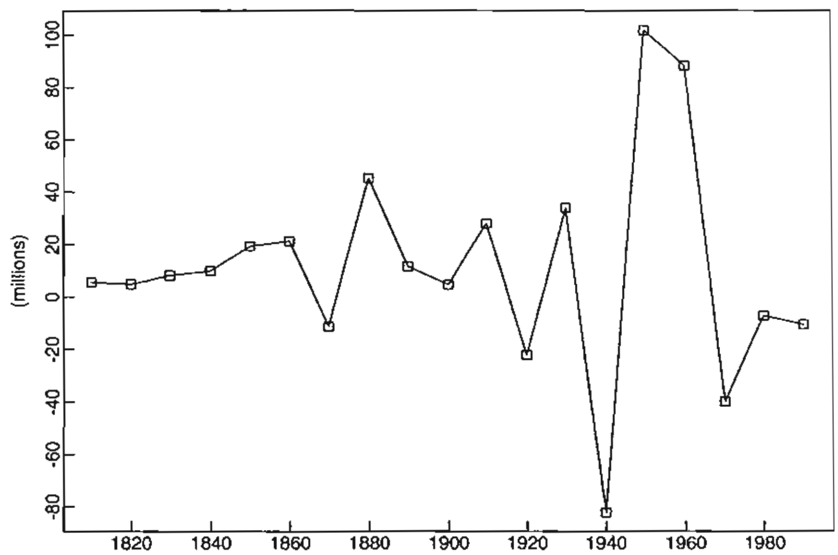

Figure 1-23
The twice-differenced series derived from the population data of Figure 1.5.

Differencing can remove <u>BOTH</u> TREND AND SEASONALITY:

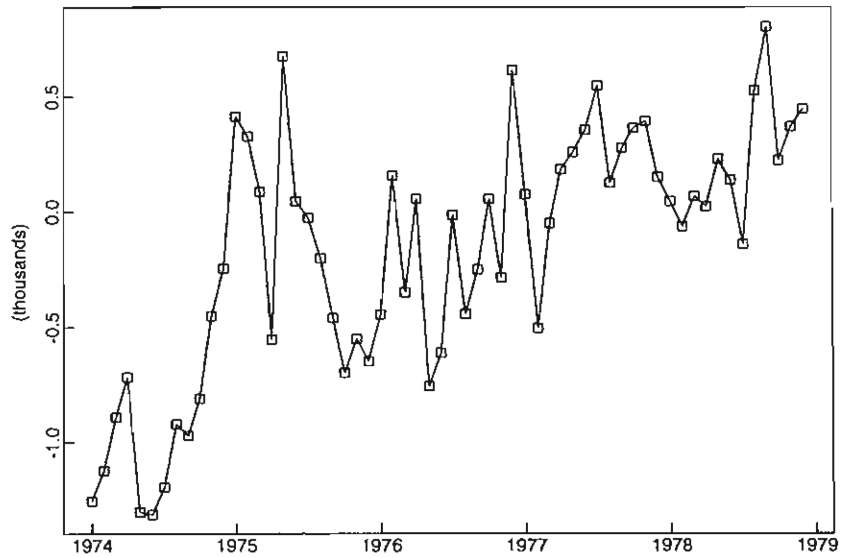

Figure 1-26
The differenced series $\{\nabla_{12}x_t, t = 13, \ldots, 72\}$ derived from the monthly accidental deaths $\{x_t, t = 1, \ldots, 72\}$.

of $\nabla\nabla_{12}x_t$, $14 \le t \le 72$, shown in Figure 1.27, which has no apparent trend or seasonal component. In Chapter 5 we shall show that this doubly differenced series can in fact be well represented by a stationary time series model. □

In this section we have discussed a variety of methods for estimating and/or removing trend and seasonality. The particular method chosen for any given data set will depend on a number of factors including whether or not estimates of the

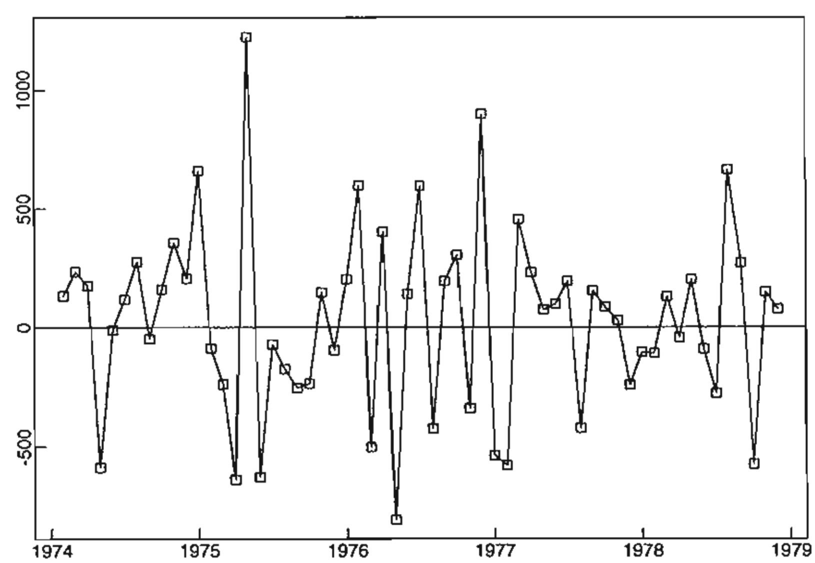

Figure 1-27
The differenced series $\{\nabla\nabla_{12}x_t, t = 14, \ldots, 72\}$ derived from the monthly accidental deaths $\{x_t, t = 1, \ldots, 72\}$.

AFTER REMOVING THE TREND AND SEASONALITY WE OBTAIN A TIME SERIES WHICH HAS A CONSTANT MEAN, $\hat{Y}_t$

- it is either a white noise process with autocorrelations at all lags $h \neq 0$ equal to 0
- it is an autocorrelated process that needs to be modelled

ex. $\hat{Y}_t = X_t - 10.202 - (-0.0242) \cdot t$

is still autocorrelated:

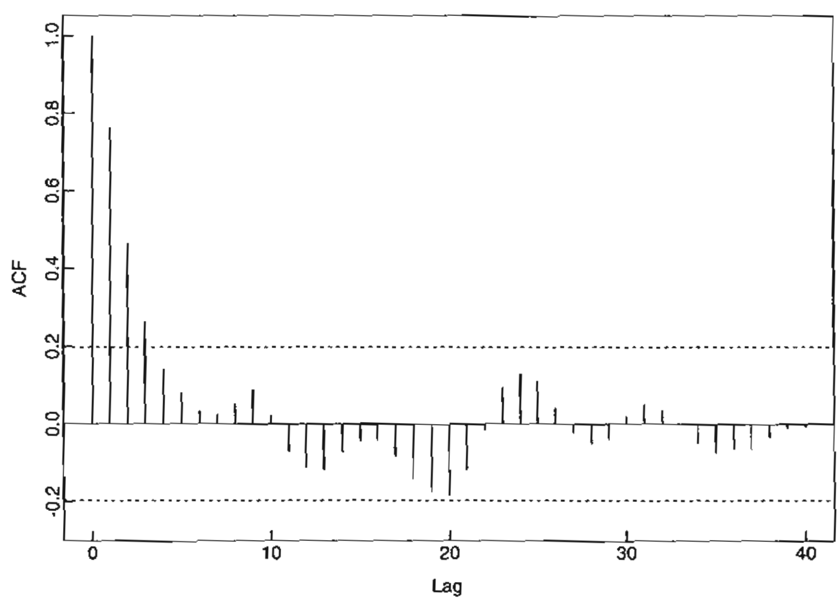

Figure 1-15
The sample autocorrelation function for the Lake Huron residuals of Figure 1.10 showing the bounds $\pm 1.96/\sqrt{n}$.

We can model it as an $\underline{AR(1)}$ process

$$Y_t = \rho Y_{t-1} + e_t$$

where e_t is a white noise.

to do that we regress by OLS $\hat{y}_t$ on $\hat{Y}_{t-1}$, and obtain

$$\hat{\hat{Y}}_t = 0.791 \hat{\hat{Y}}_{t-1} + \hat{e}_t$$

Next, we check if the ACF of $\hat{e}_t$ is zero at all lags $h \neq 0$. If not, we can try to fit an $\underline{AR(2)}$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

again by regressing $\hat{Y}_t$ on $\hat{Y}_{t-1}$ and $\hat{Y}_{t-2}$ by the OLS. We obtain

$$\hat{Y}_t = 1.002 \hat{Y}_{t-1} + (-0.2834) \hat{Y}_{t-2} + \hat{e}_t$$

and check if $\hat{e}_t$ is white noise.