Autocovariance and Autocorrelation Functions

ACF

Stochastic process

A stochastic process (or a random process) is a collection of random variables ordered by time.

A stochastic process is a set of random variables $\{y_t\}$ where the index t takes values in a certain set C. In our case, this set is ordered and corresponds to moments of time (days, months, years, etc.).

For each value of t in set C (for each point in time) a random variable, y_t , is defined and the observed values of the random variables at different times form a time series. That is, a series of T data, ($y_1, \ldots, y_t$), is a sample of size one of the vector of T random variables ordered in time corresponding to the moments t = 1, . . . , T, and the observed series is considered a result or trajectory of the stochastic process.

Stationary time series

A very important type of time series is a stationary time series. A time series is said to be strictly stationary if its properties are not affected by a change in the time origin. That is, if the joint probability distribution of the observations y_t, $y_{t+1} \ldots . y_{t+n}$ is exactly the same as the joint probability distribution of the observations y_{t+k}, $y_{t+k+1} \ldots\ldots$ y_{t+k+n} then the time series is strictly stationary. When n = 0 the stationarity assumption means that the probability distribution of y_t is the same for all time periods and can be written as f(y).

Autocovariance and Autocorrelation Functions

If a time series is stationary this means that the joint probability distribution of any two observations, say, $y_t$and y_{t+k}, is the same for any two time periods t and t + k that are separated by the same interval k. Useful information about this joint distribution, and hence about the nature of the time series, can be obtained by plotting a scatter diagram of all of the data pairs y_t , y_{t+k} that are separated by the same interval k. The interval k is called the lag.

Autocovariance and Autocorrelation Functions

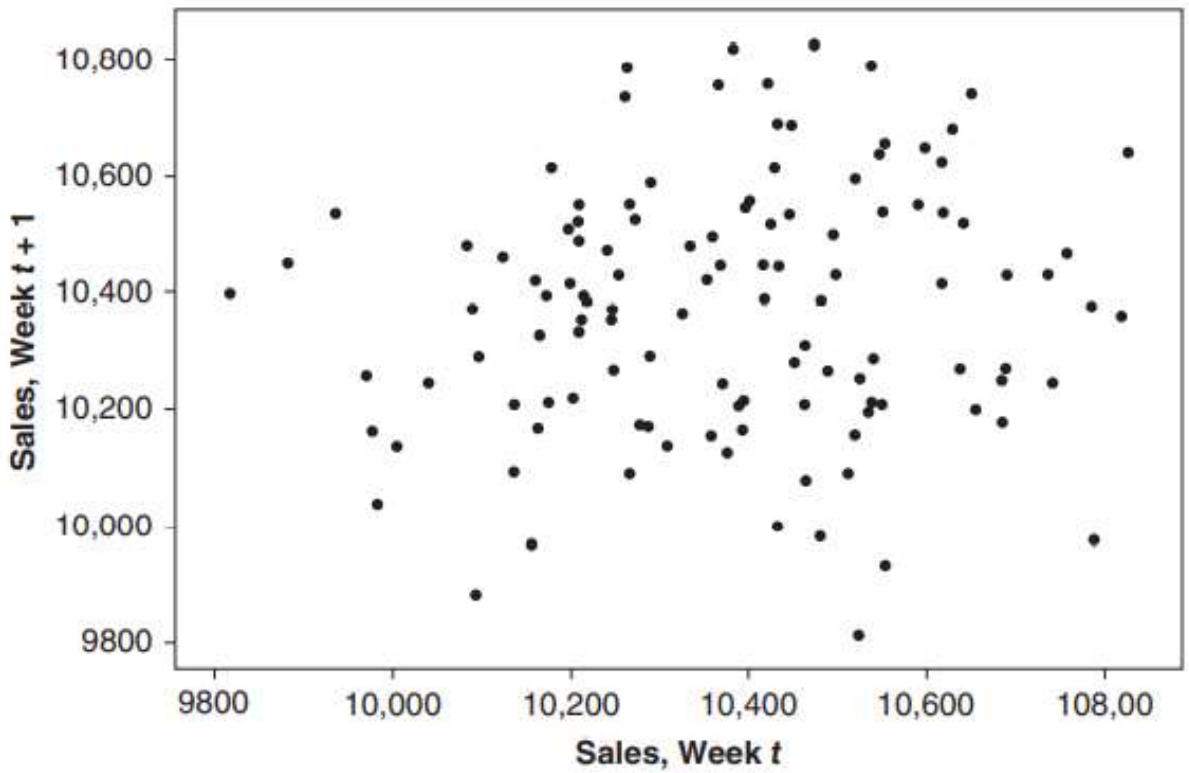

FIGURE 2.10 Scatter diagram of pharmaceutical product sales at lag $k = 1$.

Autocovariance and Autocorrelation Functions

Figure 2.10 is a scatter diagram for the pharmaceutical product sales for lag k = 1 and Figure 2.11 is a scatter diagram for the chemical viscosity readings for lag k = 1. Both scatter diagrams were constructed by plotting $y_{t+1}\ versus\ y_t$. Figure 2.10 exhibits little structure; the plotted pairs of adjacent observations y_t, y_{t+1} seem to be uncorrelated. That is, the value of y in the current period does not provide any useful information about the value of y that will be observed in the next period. A different story is revealed in Figure 2.11, where we observe that the

Autocovariance and Autocorrelation Functions

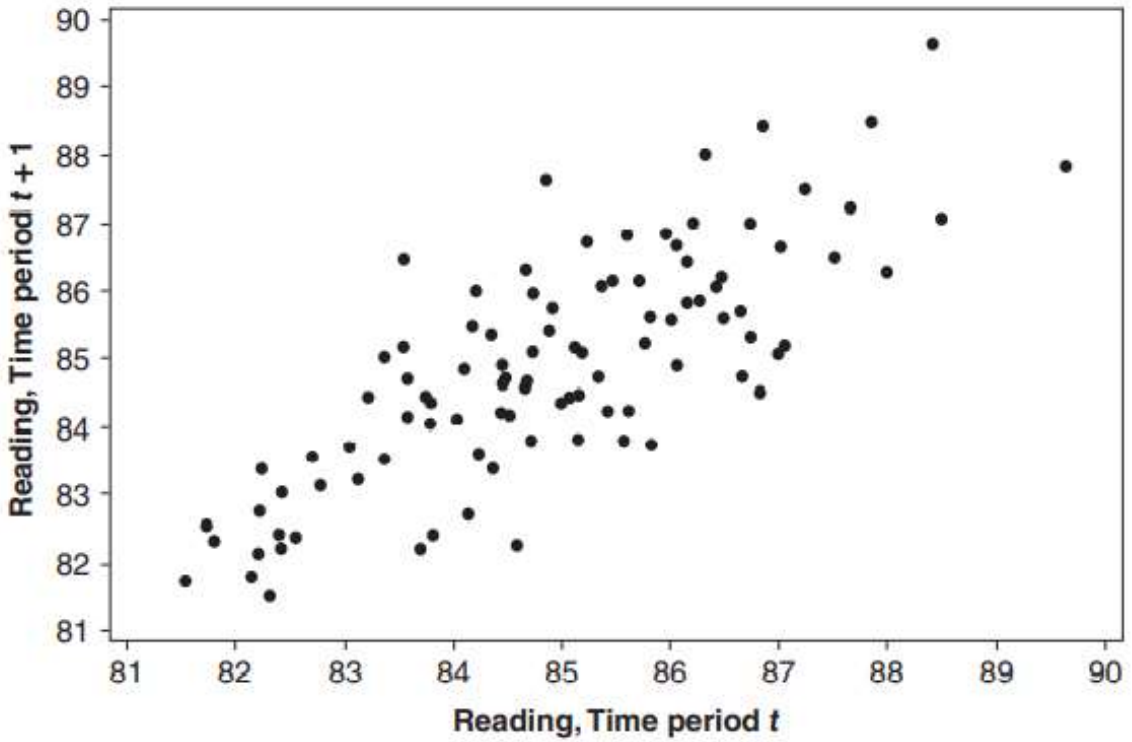

FIGURE 2.11 Scatter diagram of chemical viscosity readings at lag $k = 1$.

pairs of adjacent observations y_{t+1}, y_t are positively correlated. That is, a small value of y tends to be followed in the next time period by another small value of y, and a large value of y tends to be followed immediately by another large value of y. Note from inspection of Figures 2.10 and 2.11 that the behavior inferred from inspection of the scatter diagrams is reflected in the observed time series.

Autocovariance and Autocorrelation Functions

The covariance between y_t and its value at another time period, say, y_{t+k} is called the **autocovariance** at lag k, defined by

$$\gamma_k = \text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]. \tag{2.10}$$

The collection of the values of $\gamma_k, k = 0, 1, 2, \ldots$ is called the **autocovariance function**. Note that the autocovariance at lag $k = 0$ is just the variance of the time series; that is, $\gamma_0 = \sigma_y^2$,which is constant for a stationary time series. The **autocorrelation coefficient** at lag k for a stationary time series is

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}} = \frac{\text{Cov}(y_t, y_{t+k})}{\text{Var}(y_t)} = \frac{\gamma_k}{\gamma_0}. \tag{2.11}$$

Autocovariance and Autocorrelation Functions

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}} = \frac{\text{Cov}(y_t, y_{t+k})}{\text{Var}(y_t)} = \frac{\gamma_k}{\gamma_0}. \quad (2.11)$$

The collection of the values of ρ_k, $k = 0, 1, 2, \ldots$ is called the **autocorrelation function (ACF)**. Note that by definition $\rho_0 = 1$. Also, the ACF is independent of the scale of measurement of the time series, so it is a dimensionless quantity. Furthermore, $\rho_k = \rho_{-k}$; that is, the ACF is **symmetric** around zero, so it is only necessary to compute the positive (or negative) half.

Autocovariance and Autocorrelation Functions

It is necessary to estimate the autocovariance and ACFs from a time series of finite length, say, $y_1, y_2, \ldots, y_T$. The usual estimate of the autocovariance function is

$$c_k = \hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 0, 1, 2, \ldots, K \qquad (2.12)$$

and the ACF is estimated by the **sample autocorrelation function** (or **sample ACF**)

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}, \quad k = 0, 1, \ldots, K \qquad (2.13)$$

Autocovariance and Autocorrelation Functions

A good general rule of thumb is that at least 50 observations are required to give a reliable estimate of the ACF, and the individual sample autocorrelations should be calculated up to lag K, where K is about $T/4$.

Often we will need to determine if the autocorrelation coefficient at a particular lag is zero. This can be done by comparing the sample autocorrelation coefficient at lag k, r_k, to its standard error. If we make the assumption that the observations are uncorrelated, that is, $\rho_k = 0$ for all k, then the variance of the sample autocorrelation coefficient is

$$\text{Var}(r_k) \cong \frac{1}{T} \tag{2.14}$$

and the standard error is

$$se(r_k) \cong \frac{1}{\sqrt{T}} \tag{2.15}$$