IDENTIFICATION OF ARMA MODELS

A stationary stochastic process can be characterised, equivalently, by its autocovariance function or its partial autocovariance function.

It can also be characterised by is spectral density function, which is the Fourier transform of the autocovariances $\{\gamma_\tau; \tau = 0, \pm 1, \pm 2, \ldots\}$:

$$f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau \cos(\omega\tau) = \gamma_0 + 2\sum_{\tau=1}^{\infty} \gamma_\tau \cos(\omega\tau).$$

Here, $\omega \in [0, \pi]$ is an angular velocity, or frequency value, in radians per period.

The empirical counterpart of the spectral density function is the periodogram $I(\omega_j)$, which may be defined as

$$\frac{1}{2}I(\omega_j) = \sum_{\tau=1-T}^{T-1} c_\tau \cos(\omega_j\tau) = c_0 + 2\sum_{\tau=1}^{T-1} c_\tau \cos(\omega_j\tau),$$

where $\omega_j = 2\pi j/T; j = 0, 1, \ldots, [T/2]$ are the Fourier frequencies and $\{c_\tau; \tau = 0, \pm 1, \ldots, \pm(T-1)\}$, with $c_\tau = T^{-1}\sum_{t=\tau}^{T-1}(y_t - \bar{y})(y_{t-\tau} - \bar{y})$, are the empirical autocovariances.

The Periodogram and the Autocovariances

We need to show this definition of the peridogram is equivalent to the previous definition, which was based on the following frequency decomposition of the sample variance:

$$\frac{1}{T}\sum_{t=0}^{T-1}(y_t-\bar{y})^2=\frac{1}{2}\sum_{j=0}^{[T/2]}(\alpha_j^2+\beta_j^2),$$

where

$$\alpha_j=\frac{2}{T}\sum_t y_t\cos(\omega_j t)=\frac{2}{T}\sum_t(y_t-\bar{y})\cos(\omega_j t),$$

$$\beta_j=\frac{2}{T}\sum_t y_t\sin(\omega_j t)=\frac{2}{T}\sum_t(y_t-\bar{y})\sin(\omega_j t).$$

Substituting these into the term $T(\alpha_j^2+\beta_j^2)/2$ gives the periodogram

$$I(\omega_j)=\frac{2}{T}\left[\left\{\sum_{t=0}^{T-1}\cos(\omega_j t)(y_t-\bar{y})\right\}^2+\left\{\sum_{t=0}^{T-1}\sin(\omega_j t)(y_t-\bar{y})\right\}^2\right].$$

The quadratic terms may be expanded to give

$$\begin{aligned}I(\omega_j) =& \frac{2}{T}\left\{\sum_t\sum_s \cos(\omega_j t)\cos(\omega_j s)(y_t-\bar{y})(y_s-\bar{y})\right\}\\&+\frac{2}{T}\left\{\sum_t\sum_s \sin(\omega_j t)\sin(\omega_j s)(y_t-\bar{y})(y_s-\bar{y})\right\},\end{aligned}$$

Since $\cos(A)\cos(B)+\sin(A)\sin(B)=\cos(A-B)$, this can be written as

$$I(\omega_j) = \frac{2}{T}\left\{\sum_t\sum_s \cos(\omega_j[t-s])(y_t-\bar{y})(y_s-\bar{y})\right\}$$

On defining $\tau = t-s$ and writing $c_\tau = \sum_t (y_t-\bar{y})(y_{t-\tau}-\bar{y})/T$, we can reduce the latter expression to

$$I(\omega_j) = 2\sum_{\tau=1-T}^{T-1}\cos(\omega_j\tau)c_\tau,$$

which is a Fourier transform of the empirical autocovariances.

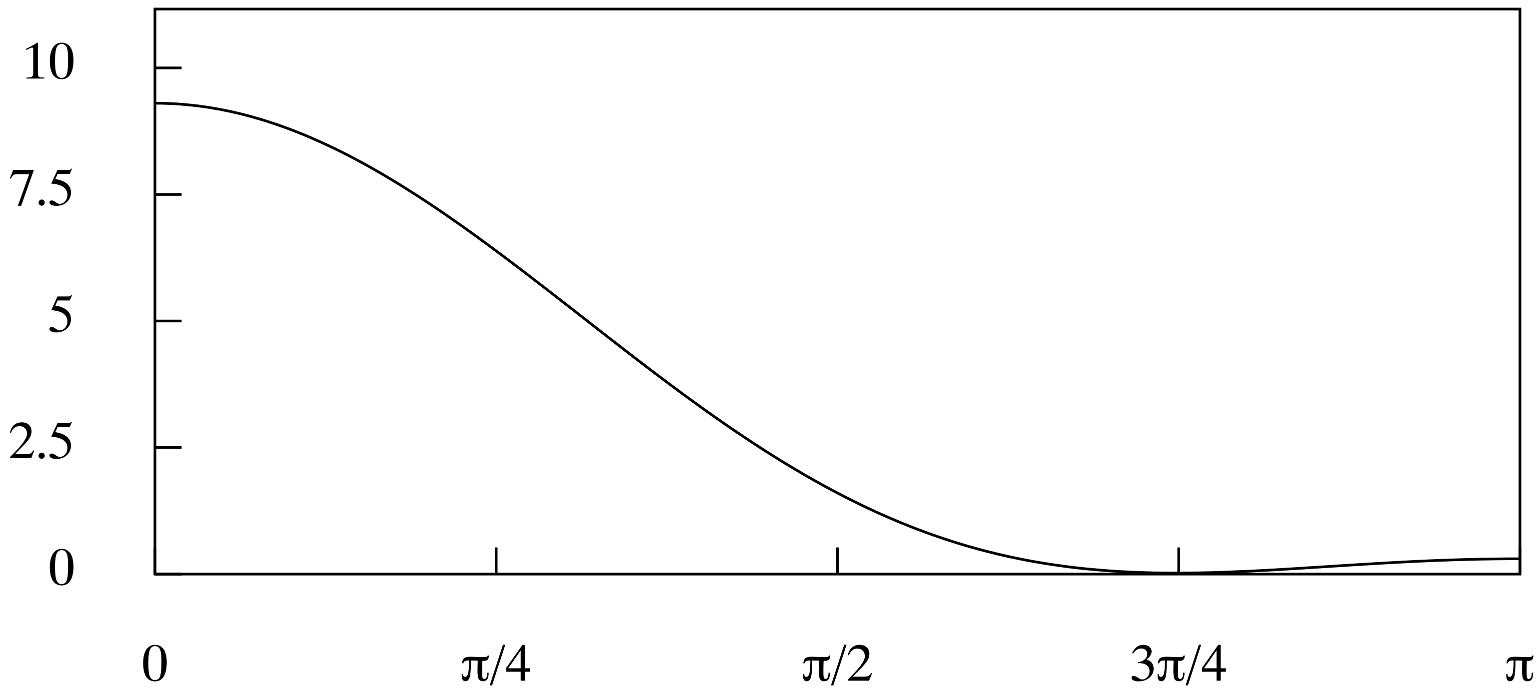

Figure 1. The spectral density function of an MA(2) process $y(t) = (1 + 1.250L + 0.800L^2)\varepsilon(t)$.

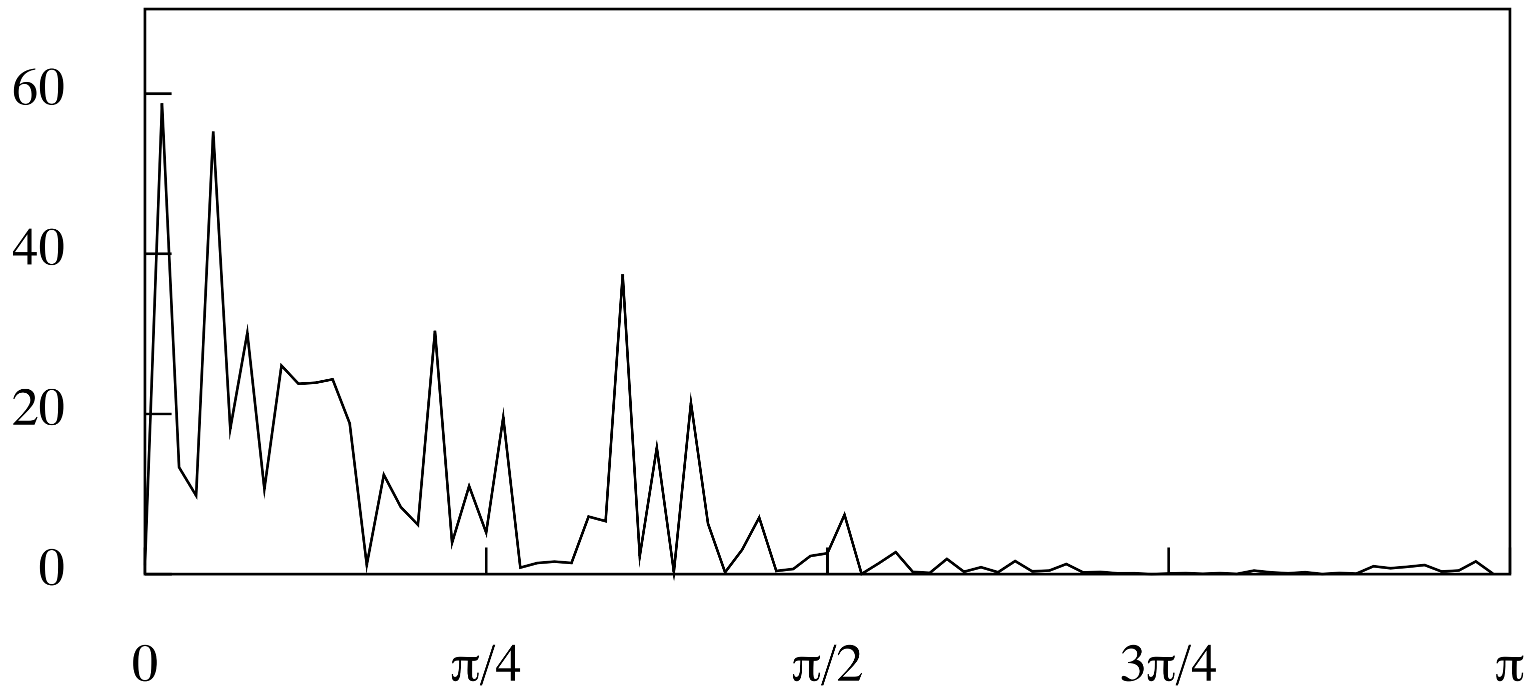

Figure 2. The graph of a periodogram calculated from 160 observations on a simulated series generated by an MA(2) process $y(t) = (1 + 1.250L + 0.800L^2)\varepsilon(t)$.

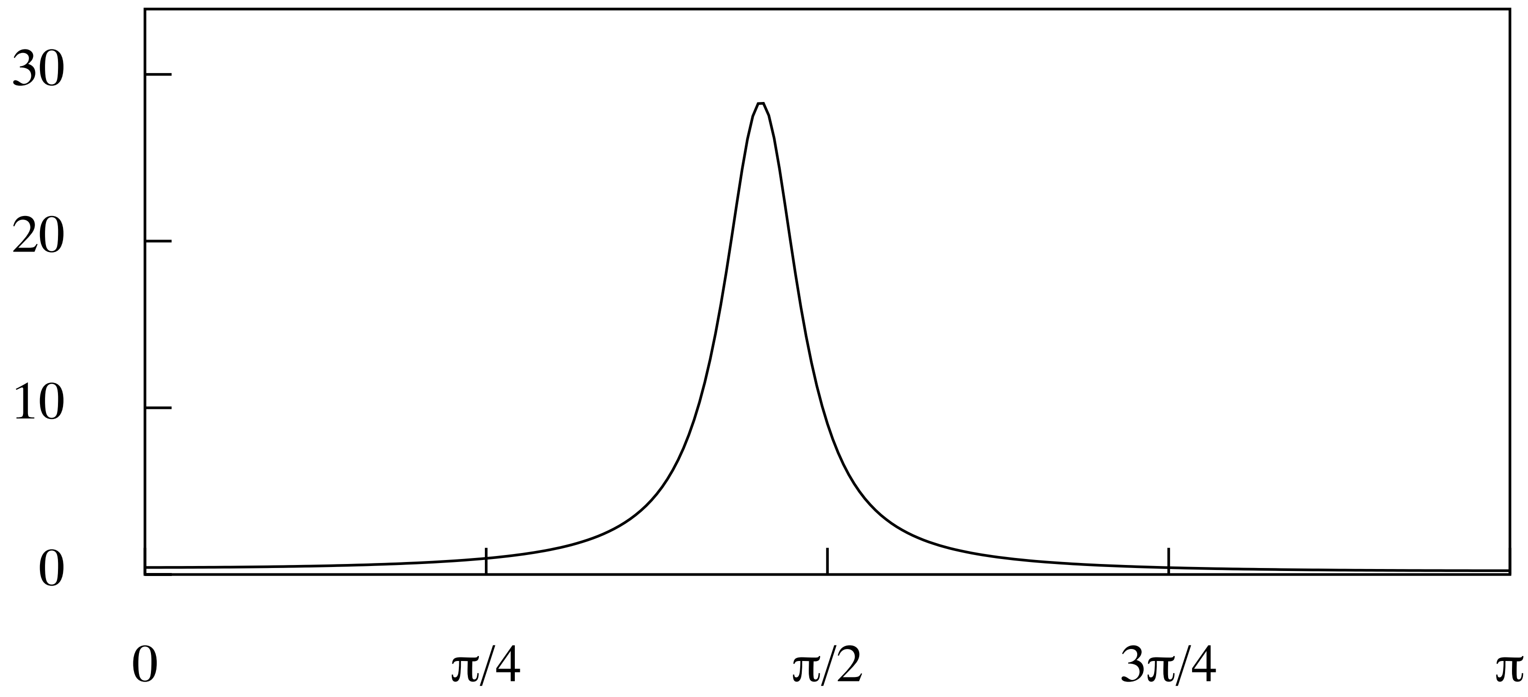

Figure 3. The spectral density function of an AR(2) process $(1 - 0.273L + 0.810L^2)y(t) = \varepsilon(t)$.

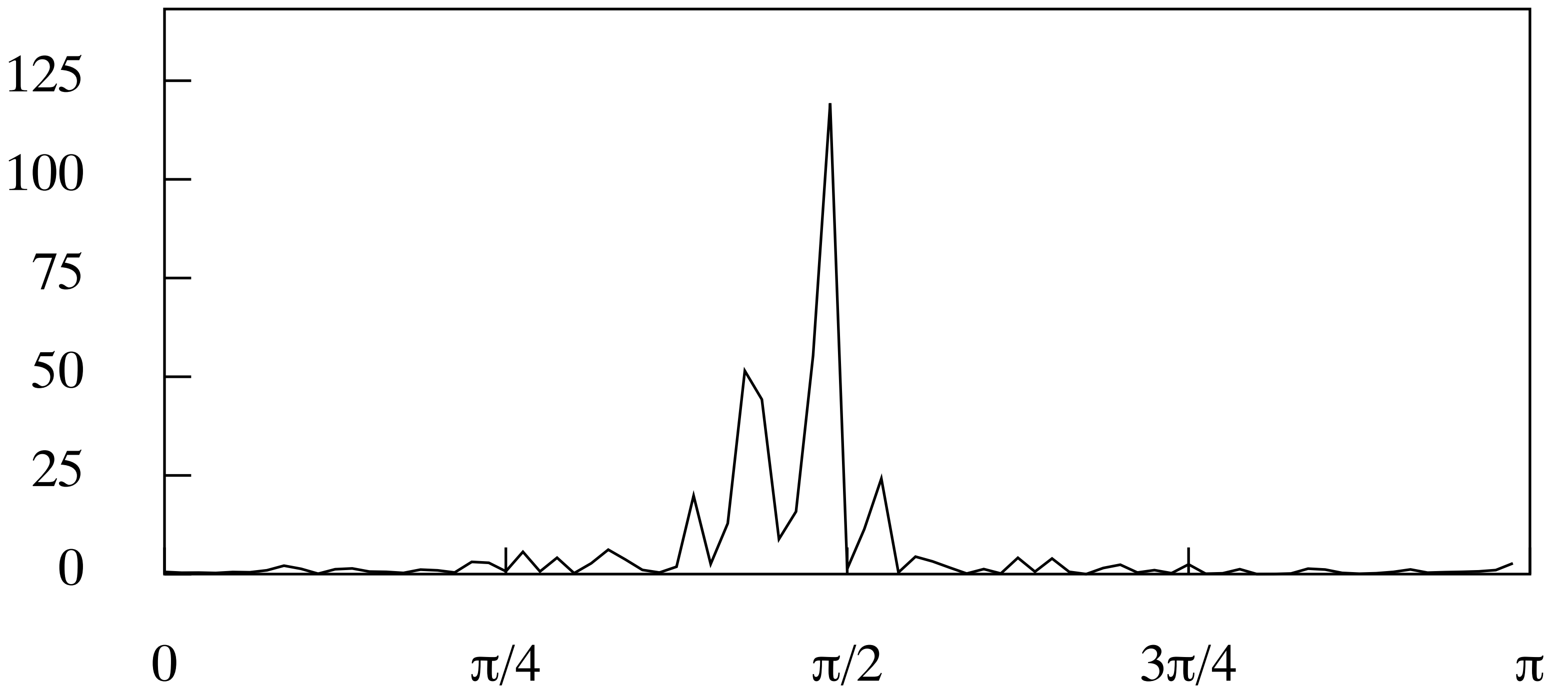

Figure 4. The graph of a periodogram calculated from 160 observations on a simulated series generated by an AR(2) process $(1 - 0.273L + 0.810L^2)y(t) = \varepsilon(t)$.

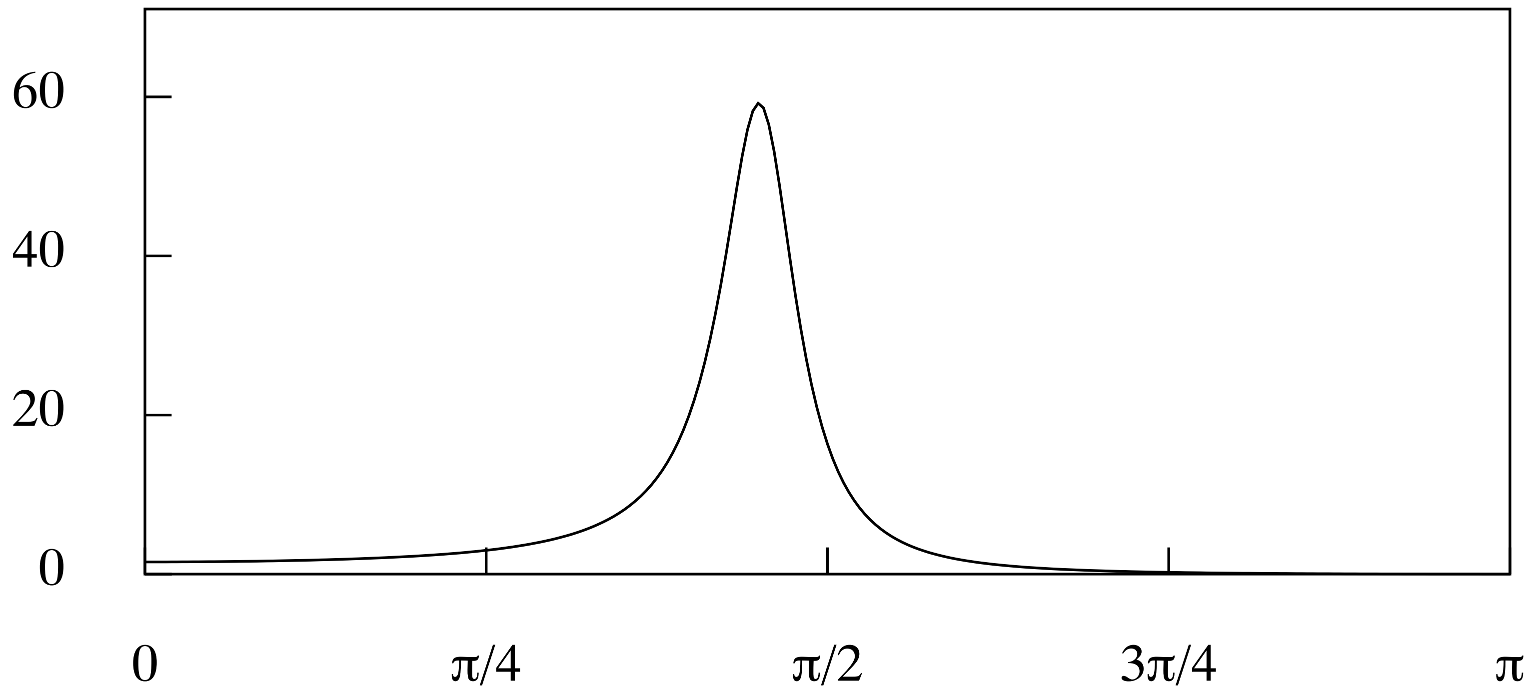

Figure 5. The spectral density function of an ARMA(2, 1) process $(1 - 0.273L + 0.810L^2)y(t) = (1 + 0.900L)\varepsilon(t)$.

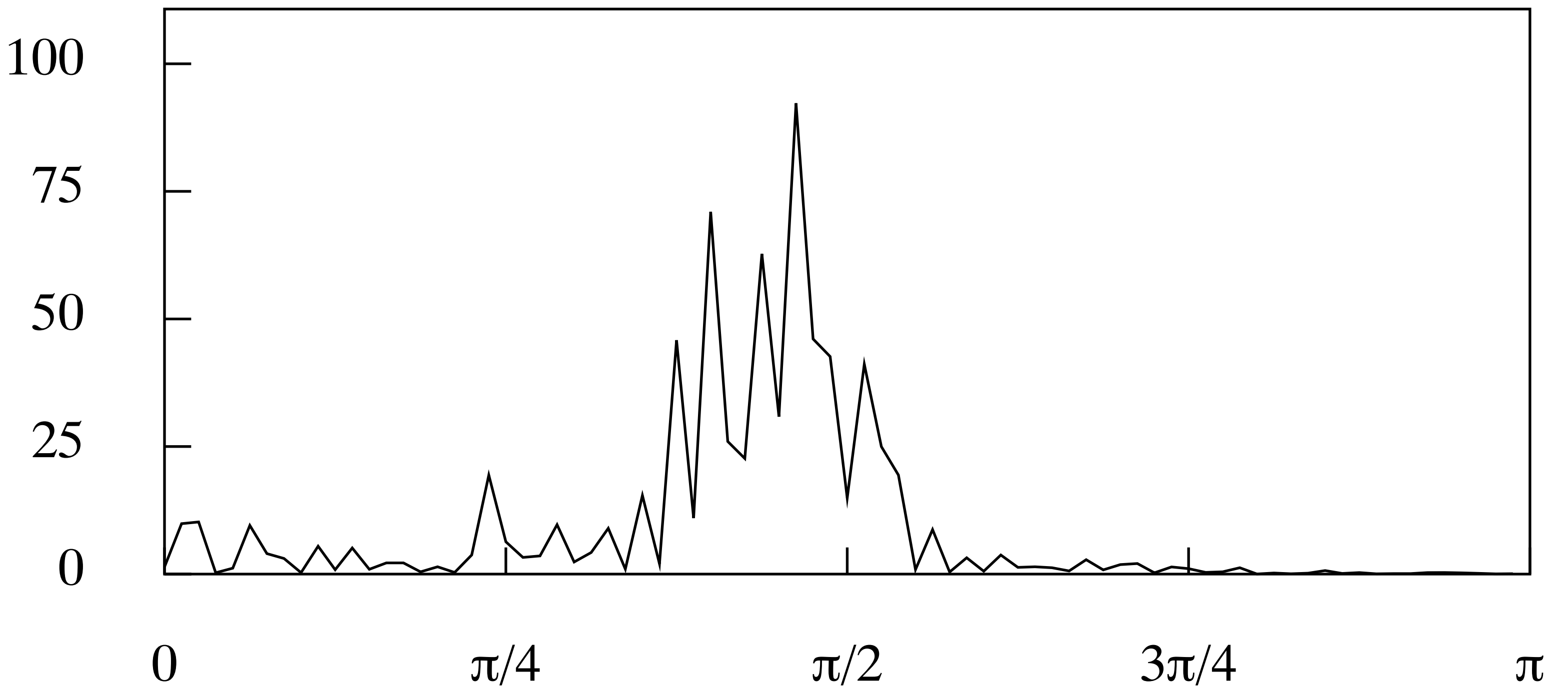

Figure 6. The graph of a periodogram calculated from 160 observations on a simulated series generated by an ARMA(2, 1) process $(1 - 0.273L + 0.810L^2)y(t) = (1 + 0.900L)\varepsilon(t)$.

The Methodology of Box and Jenkins

Box and Jenkins proposed to use the autocorrelation and partial autocorrelation functions for identifying the orders of ARMA models. They paid little attention to the periodogram.

Autocorrelation function (ACF). Given a sample $y_0, y_1, \ldots, y_{T-1}$ of T observations, the sample autocorrelation function $\{r_\tau\}$ is the sequence

$$r_\tau = c_\tau / c_0, \quad \tau = 0, 1, \ldots,$$

where $c_\tau = T^{-1} \sum (y_t - \bar{y})(y_{t-\tau} - \bar{y})$ is the empirical autocovariance at lag τ and c_0 is the sample variance.

As the lag increases, the number of observations comprised in the empirical autocovariances diminishes.

Partial autocorrelation function (PACF). The sample partial autocorrelation function $\{p_\tau\}$ gives the correlation between the two sets of residuals obtained from regressing the elements y_t and $y_{t-\tau}$ on the set of intervening values $y_{t-1}, y_{t-2}, \ldots, y_{t-\tau+1}$. The partial autocorrelation measures the dependence between y_t and $y_{t-\tau}$ after the effect of the intervening values has been removed.

Reduction to Stationarity.

The first step is to examine the plot of the data to judge whether or not the process is stationary. A trend can be removed by fitting a parametric curve or a spline function to create a stationary sequence of residuals to which an ARMA model can be applied.

Box and Jenkins believed that many empirical series can be modelled by taking a sufficient number of differences to make it stationary. Thus, the process might be modelled by the ARIMA(p, d, q) equation

$$\alpha(L)\nabla^d y(t) = \mu(L)\varepsilon(t),$$

where $\nabla^d = (I - L)^d$ is the dth power of the difference operator.

Then, $z(t) = \nabla^d y(t)$ will be described by a stationary ARMA(p, q) model. The inverse operator ∇^{-1} is the summing or integrating operator, which is why the model described an autoregressive integrated moving-average.

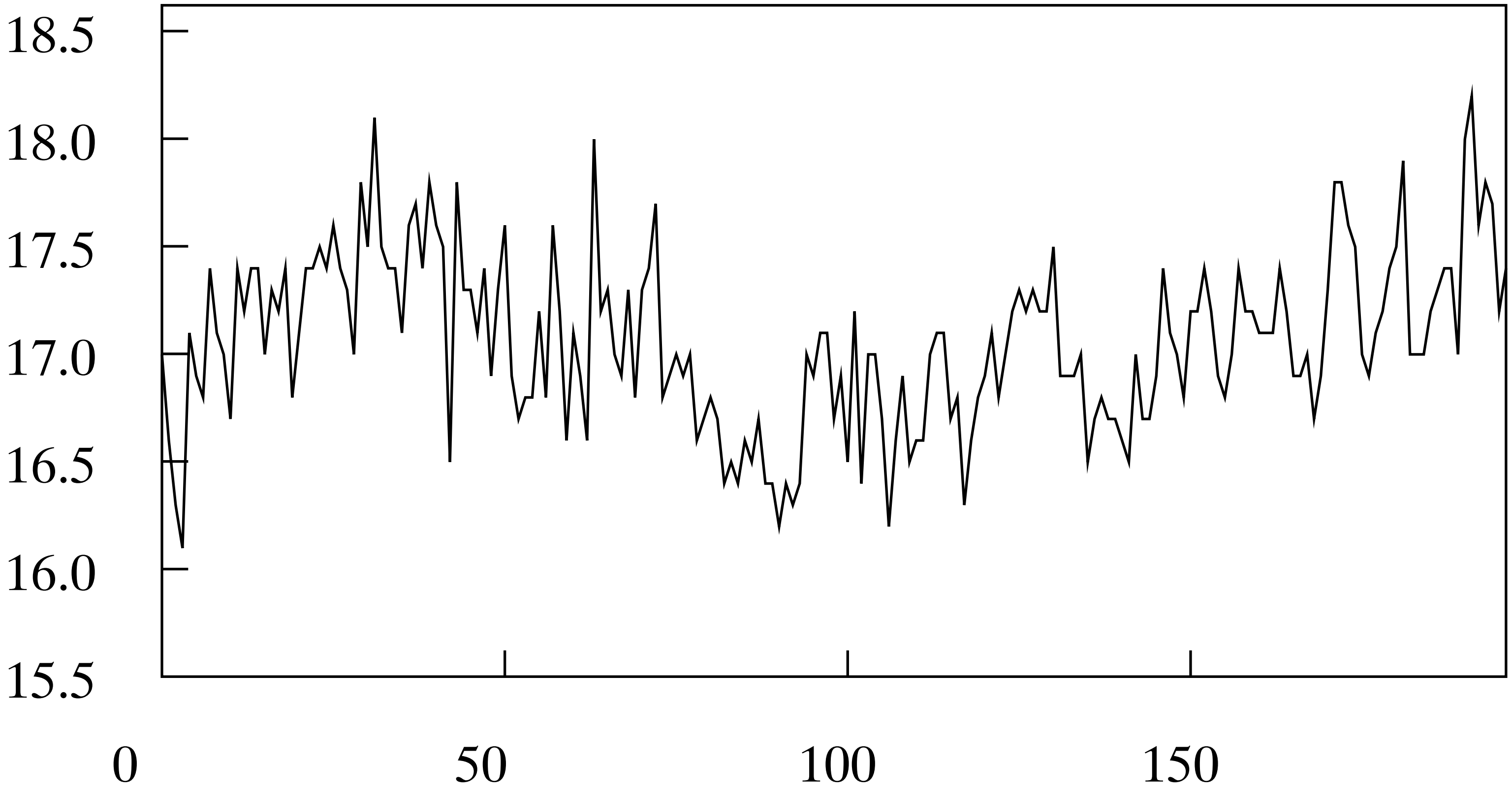

Figure 7. The plot of 197 concentration readings from a chemical process taken at 2-hour intervals.

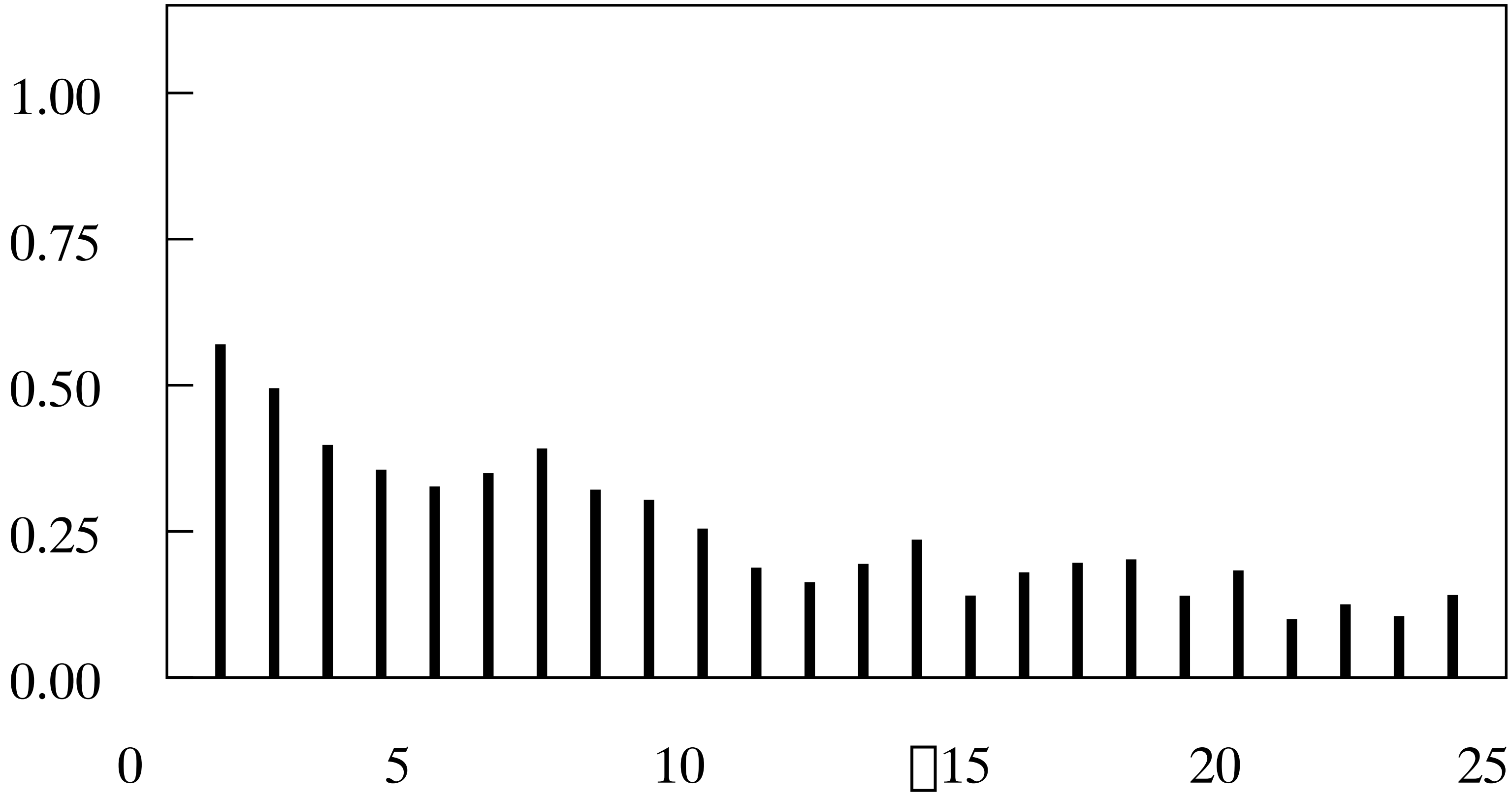

Figure 8. The autocorrelation function of the concentration readings from a chemical process.

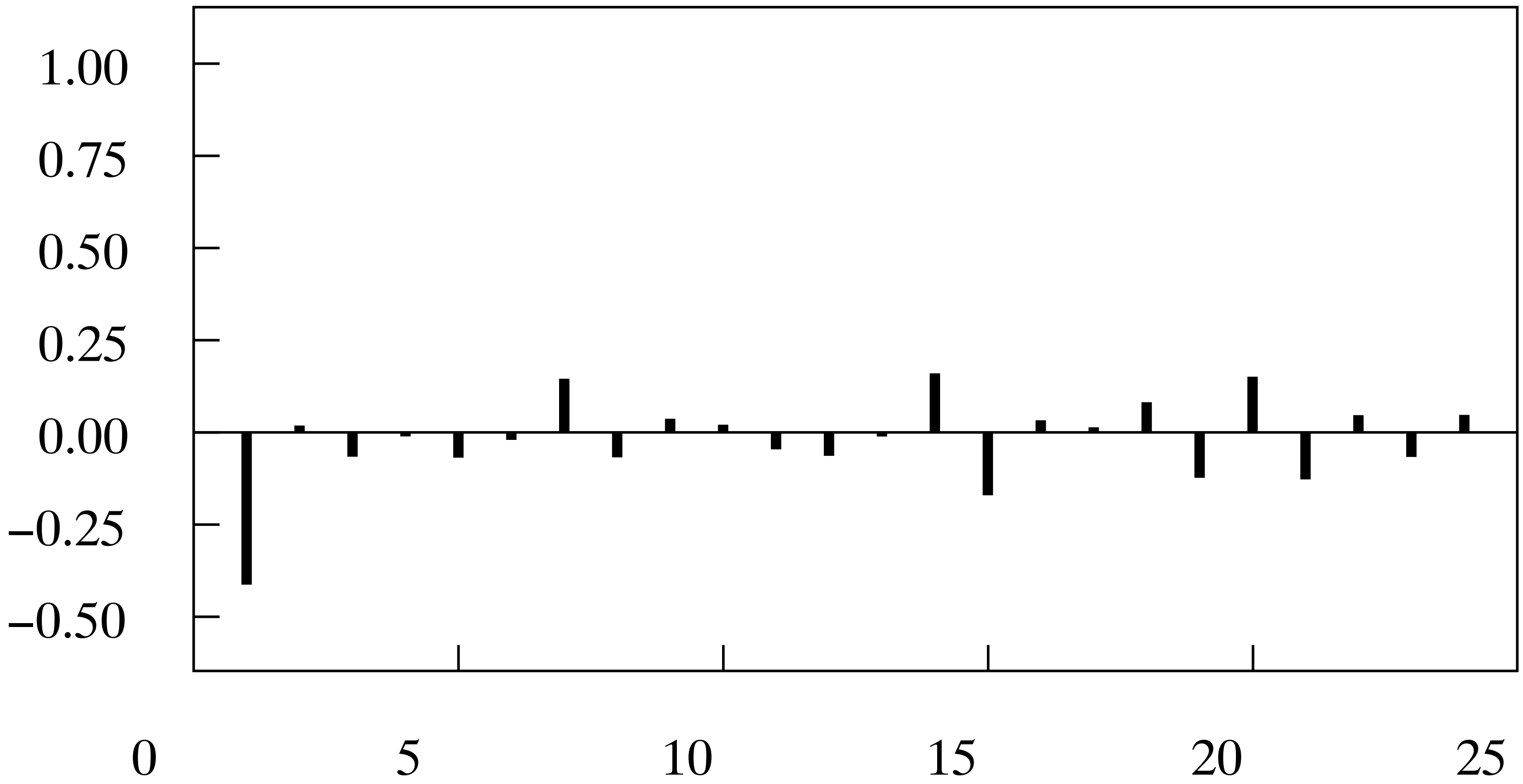

Figure 9. The autocorrelation function of the differences of the concentration readings from the chemical process.

When Stationarity has been achieved, the autocorrelation sequence of the resulting series should converge rapidly to zero as the value of the lag increases. (See Figure 9.)

The characteristics of pure autoregressive and pure moving-average process are easily spotted. Those of a mixed autoregressive moving-average model are not so easily unravelled.

Moving-average processes. The theoretical autocorrelation function $\{\rho_\tau\}$ of an M(q) process has $\rho_\tau = 0$ for all $\tau > q$. The partial autocorrelation function $\{\pi_\tau\}$ is liable to decay towards zero gradually.

To determine whether the parent autocorrelations are zero after lag q, we may use a result of Bartlett [1946] which shows that, for a sample of size T, the standard deviation of r_τ is approximately

$$(4)\qquad \frac{1}{\sqrt{T}}\left\{1 + 2(r_1^2 + r_2^2 + \cdots + r_q^2)\right\}^{1/2} \quad \text{for} \quad \tau > q.$$

A measure of the scale of the autocorrelations is provided by the limits of $\pm 1.96/\sqrt{T}$, which are the approximate 95% confidence bounds for the autocorrelations of a white-noise sequence. These bounds are represented by the dashed horizontal lines on the accompanying graphs.

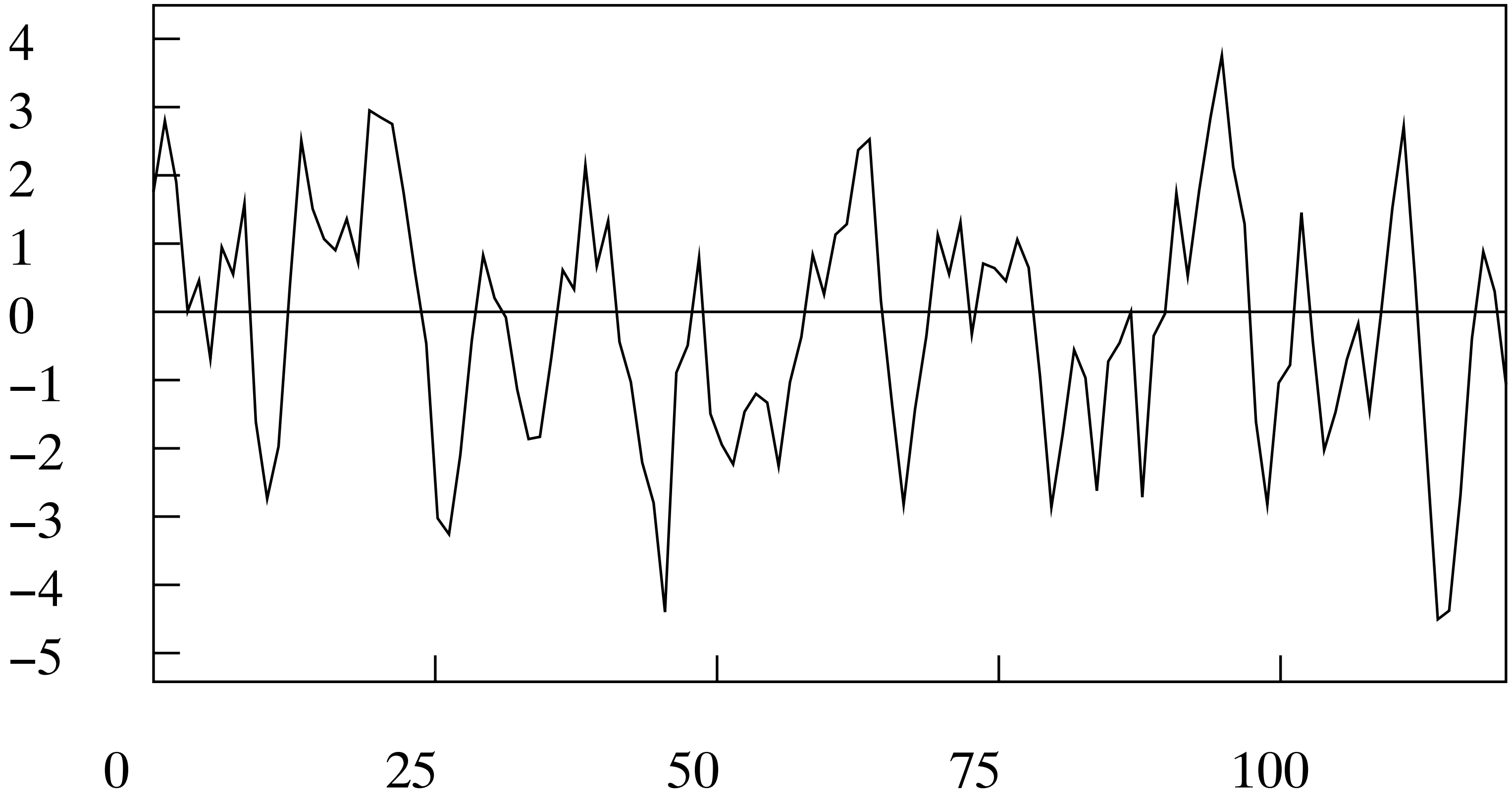

Figure 10. The graph of 120 observations on a simulated series generated by the MA(2) process $y(t) = (1 + 0.90L + 0.81L^2)\varepsilon(t)$.

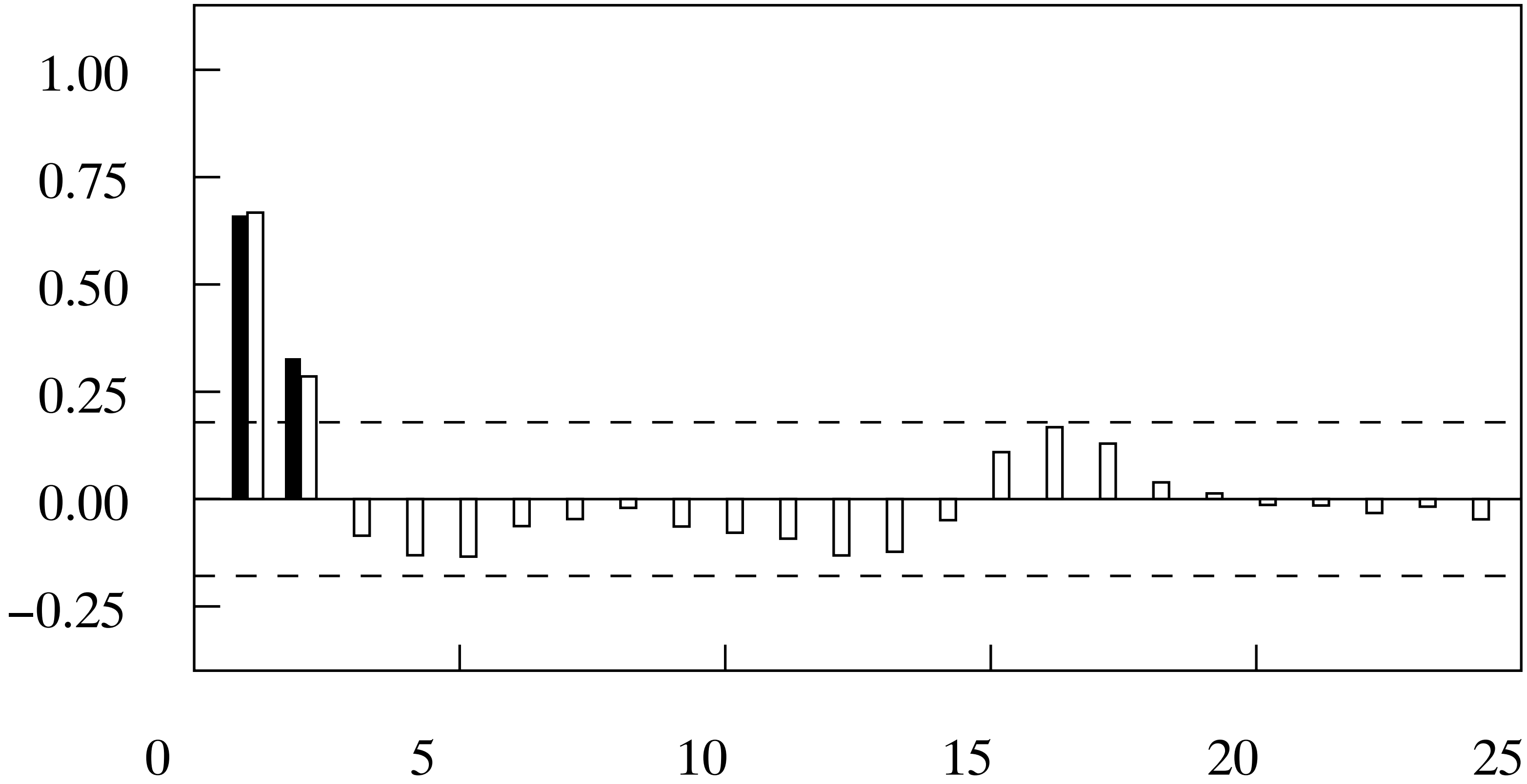

Figure 11. The theoretical autocorrelation function (ACF) of the MA(2) process $y(t) = (1 + 0.90L + 0.81L^2)\varepsilon(t)$ (the solid bars) together with its empirical counterpart, calculated from a simulated series of 120 observations.

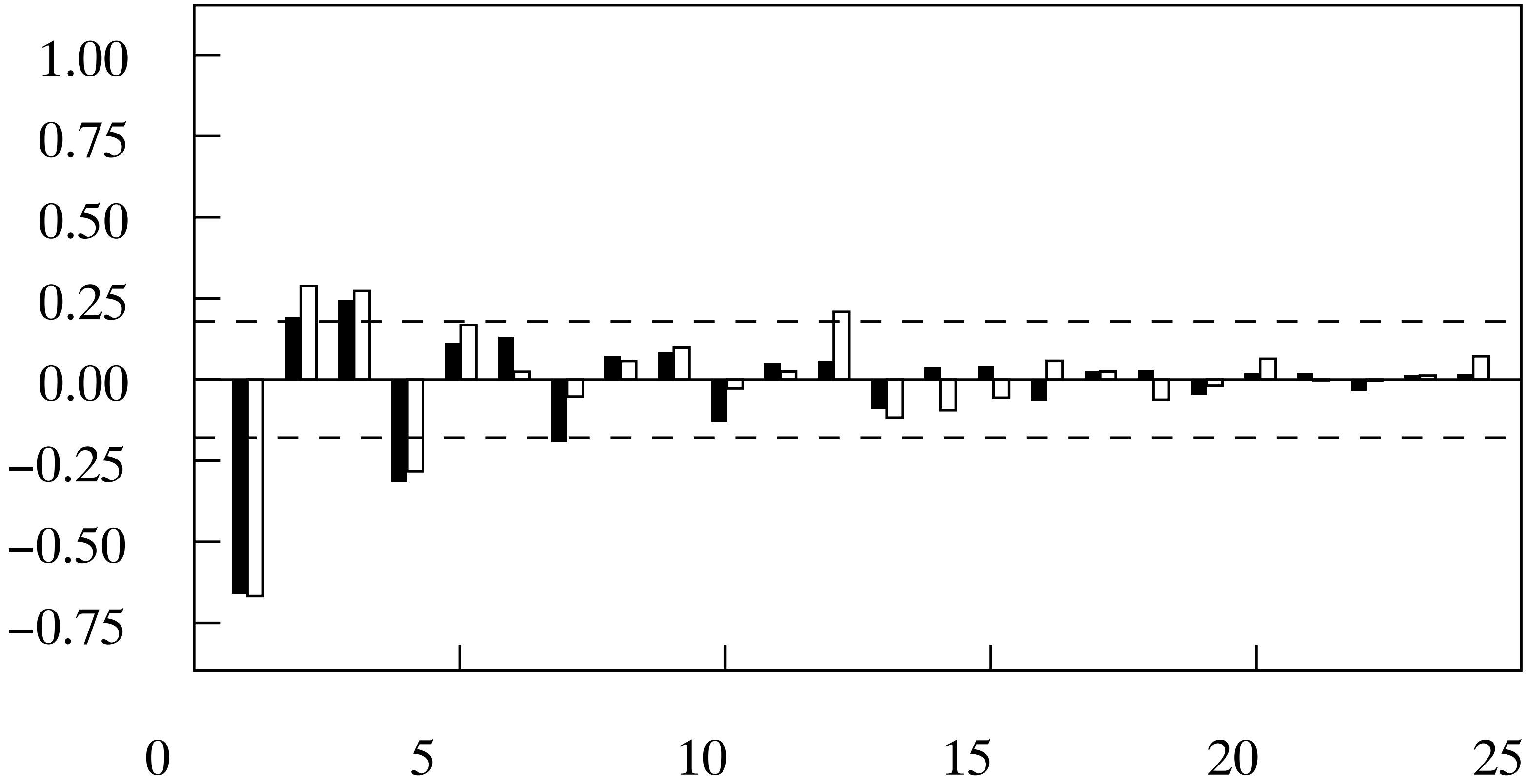

Figure 12. The theoretical partial autocorrelation function (PACF) of the MA(2) process $y(t) = (1+0.90L+0.81L^2)\varepsilon(t)$ (the solid bars) together with its empirical counterpart, calculated from a simulated series of 120 observations.

Autoregressive processes. The theoretical autocorrelation function $\{\rho_\tau\}$ of an AR(p) process obeys a homogeneous difference equation based upon the autoregressive operator $\alpha(L) = 1 + \alpha_1 L + \cdots + \alpha_p L^p$:

$$(5) \qquad \rho_\tau = -(\alpha_1 \rho_{\tau-1} + \cdots + \alpha_p \rho_{\tau-p}) \quad \text{for all} \quad \tau \geq p.$$

The autocorrelation sequence will be a mixture of damped exponential and sinusoidal functions. If the sequence is of a sinusoidal nature, then the presence of complex roots in the operator $\alpha(L)$ is indicated.

The partial autocorrelation function $\{\pi_\tau\}$ serves most clearly to identify a pure AR process. An AR(p) process has $\pi_\tau = 0$ for all $\tau > p$.

The significance of the values of the empirical partial autocorrelations is judged by the fact that, for a pth order process, their standard deviations for all lags greater that p are approximated by $1/\sqrt{T}$. The bounds of $\pm 1.96/\sqrt{T}$ are plotted on the graph of the partial autocorrelation function.

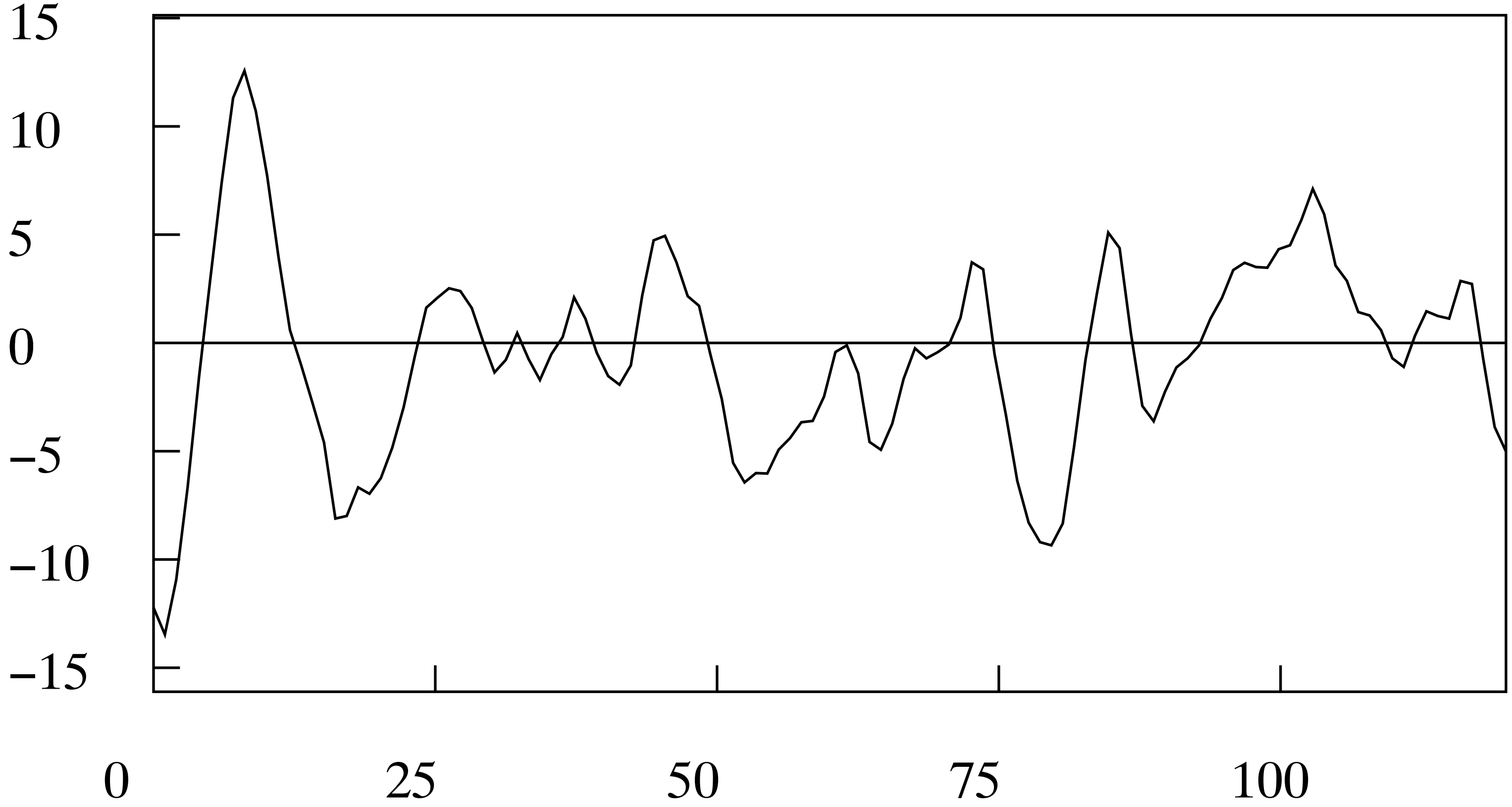

Figure 13. The graph of 120 observations on a simulated series generated by the AR(2) process $(1 - 1.69L + 0.81L^2)y(t) = \varepsilon(t)$.

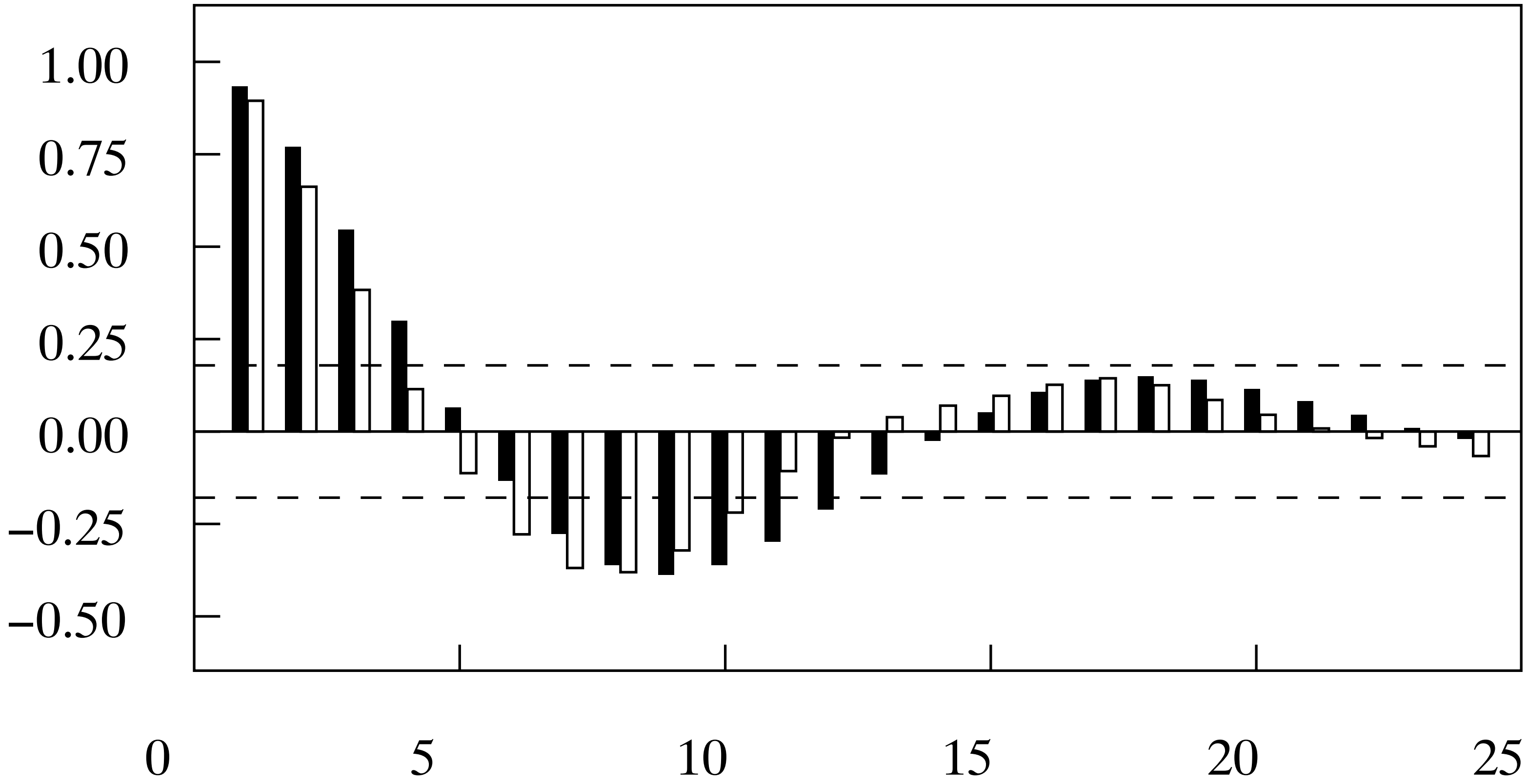

Figure 14. The theoretical autocorrelation function (ACF) of the AR(2) process $(1 - 1.69L + 0.81L^2)y(t) = \varepsilon(t)$ (the solid bars) together with its empirical counterpart, calculated from a simulated series of 120 observations.

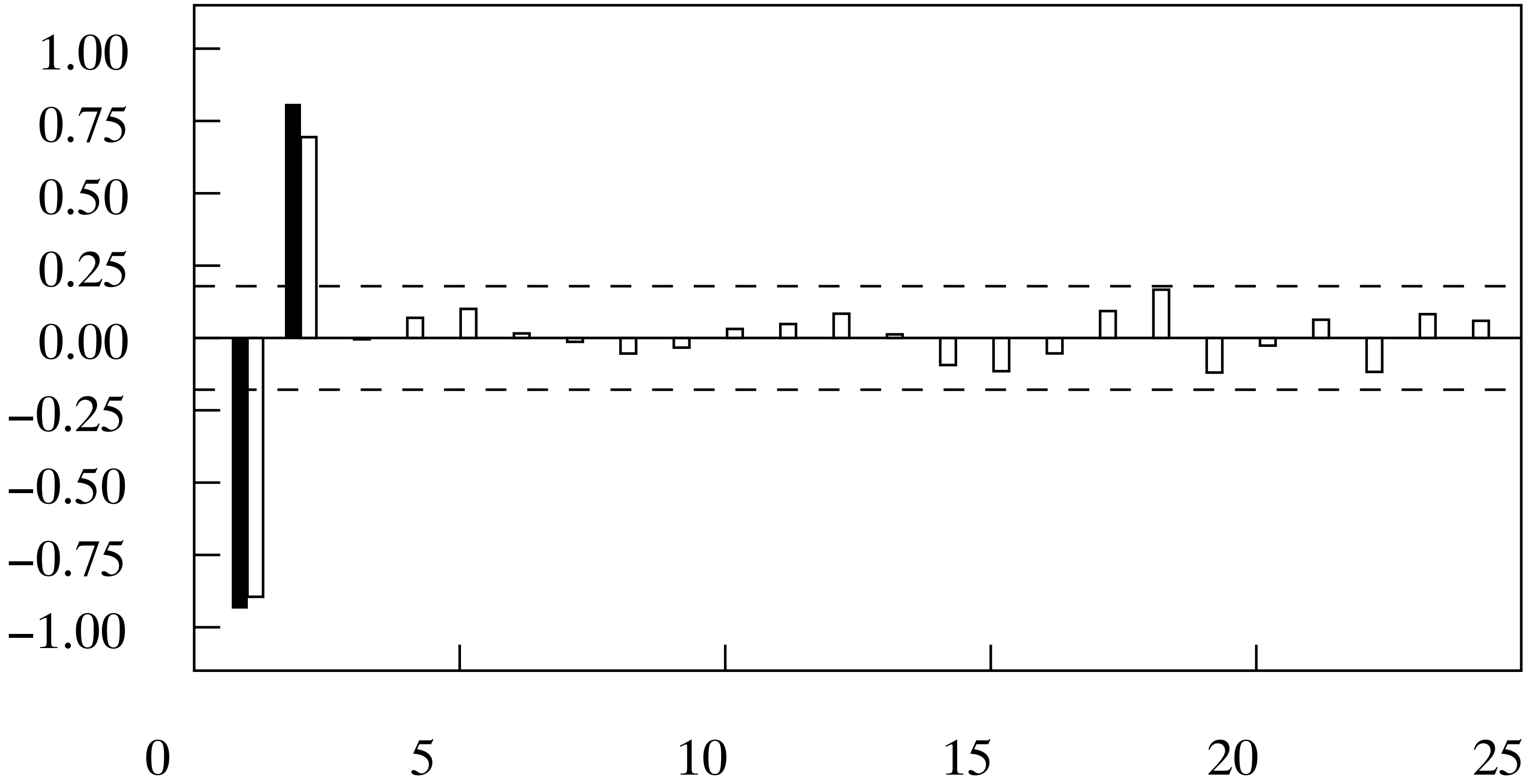

Figure 15. The theoretical partial autocorrelation function (PACF) of the AR(2) process $(1-1.69L+0.81L^2)y(t)=\varepsilon(t)$ (the solid bars) together with its empirical counterpart, calculated from a simulated series of 120 observations.

Mixed processes. Neither the theoretical autocorrelation or partial autocorrelation functions of an ARMA(p, q) process have abrupt cutoffs. The autocovariances an ARMA(p, q) process satisfy the same difference equation as that of a pure AR model for all values of $\tau > \max(p, q)$.

A rational transfer function is more effective in approximating an arbitrary impulse response than is an AR or an MA transfer function

The sum of any two mutually independent AR processes gives rise to an ARMA process. Let $y(t)$ and $z(t)$ be AR processes of orders p and r respectively described by $\alpha(L)y(t) = \varepsilon(t)$ and $\rho(L)z(t) = \eta(t)$, wherein $\varepsilon(t)$ and $\eta(t)$ are mutually independent white-noise processes. Then their sum will be

$$
\begin{aligned}
y(t) + z(t) &= \frac{\varepsilon(t)}{\alpha(L)} + \frac{\eta(t)}{\rho(L)} \\
&= \frac{\rho(L)\varepsilon(t) + \alpha(L)\eta(t)}{\alpha(L)\rho(L)} = \frac{\mu(L)\zeta(t)}{\alpha(L)\rho(L)},
\end{aligned} \tag{6}
$$

where $\mu(L)\zeta(t) = \rho(L)\varepsilon(t) + \alpha(L)\eta(t)$ constitutes a moving-average process of order $\max(p, r)$.

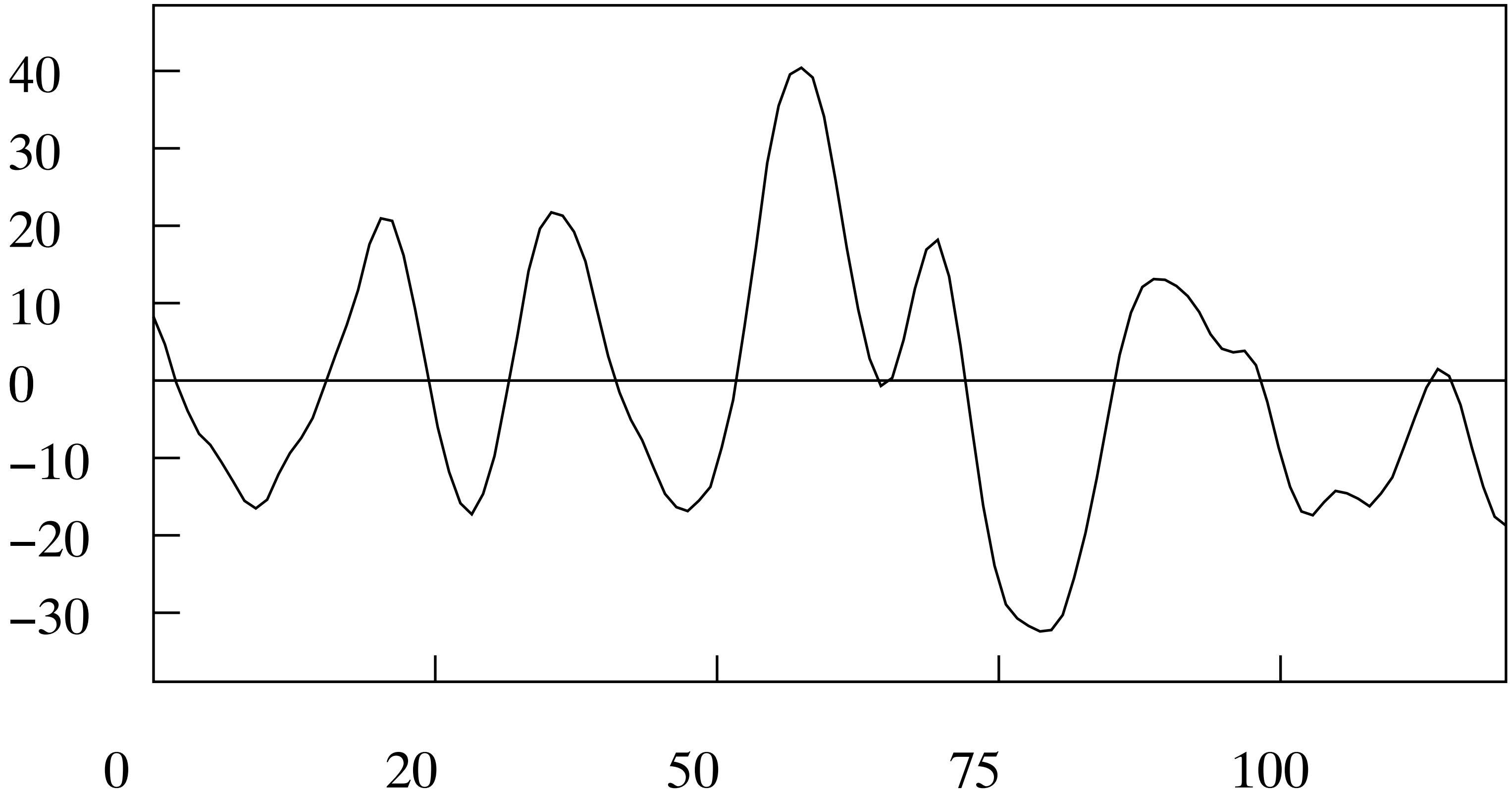

Figure 16. The graph of 120 observations on a simulated series generated by the ARMA(2, 2) process $(1-1.69L+0.81L^2)y(t) = (1+0.90L+0.81L^2)\varepsilon(t)$.

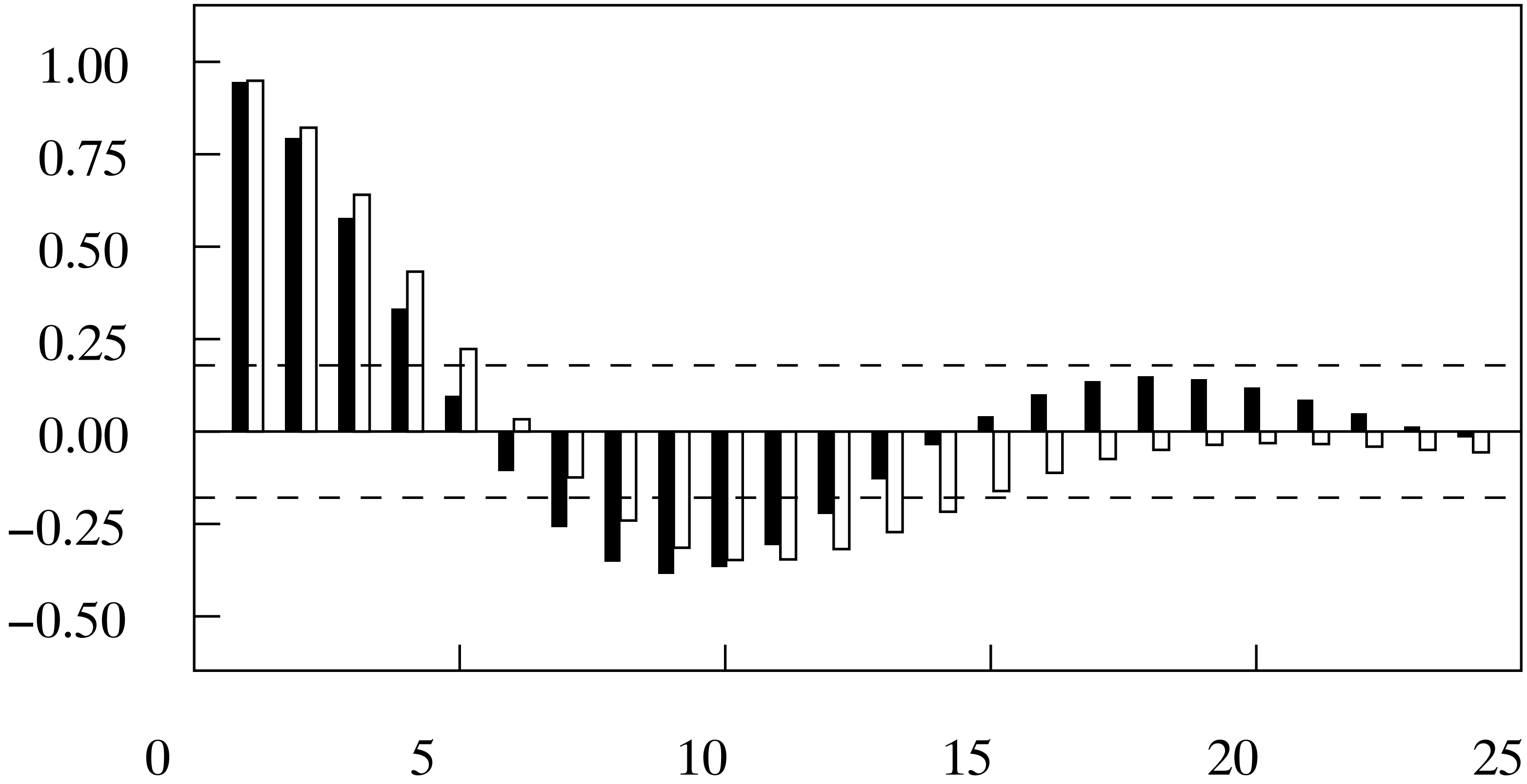

Figure 17. The theoretical autocorrelation function (ACF) of the ARMA(2, 2) process $(1 - 1.69L + 0.81L^2)y(t) = (1 + 0.90L + 0.81L^2)\varepsilon(t)$ (the solid bars) together with its empirical counterpart, calculated from a simulated series of 120 observations.

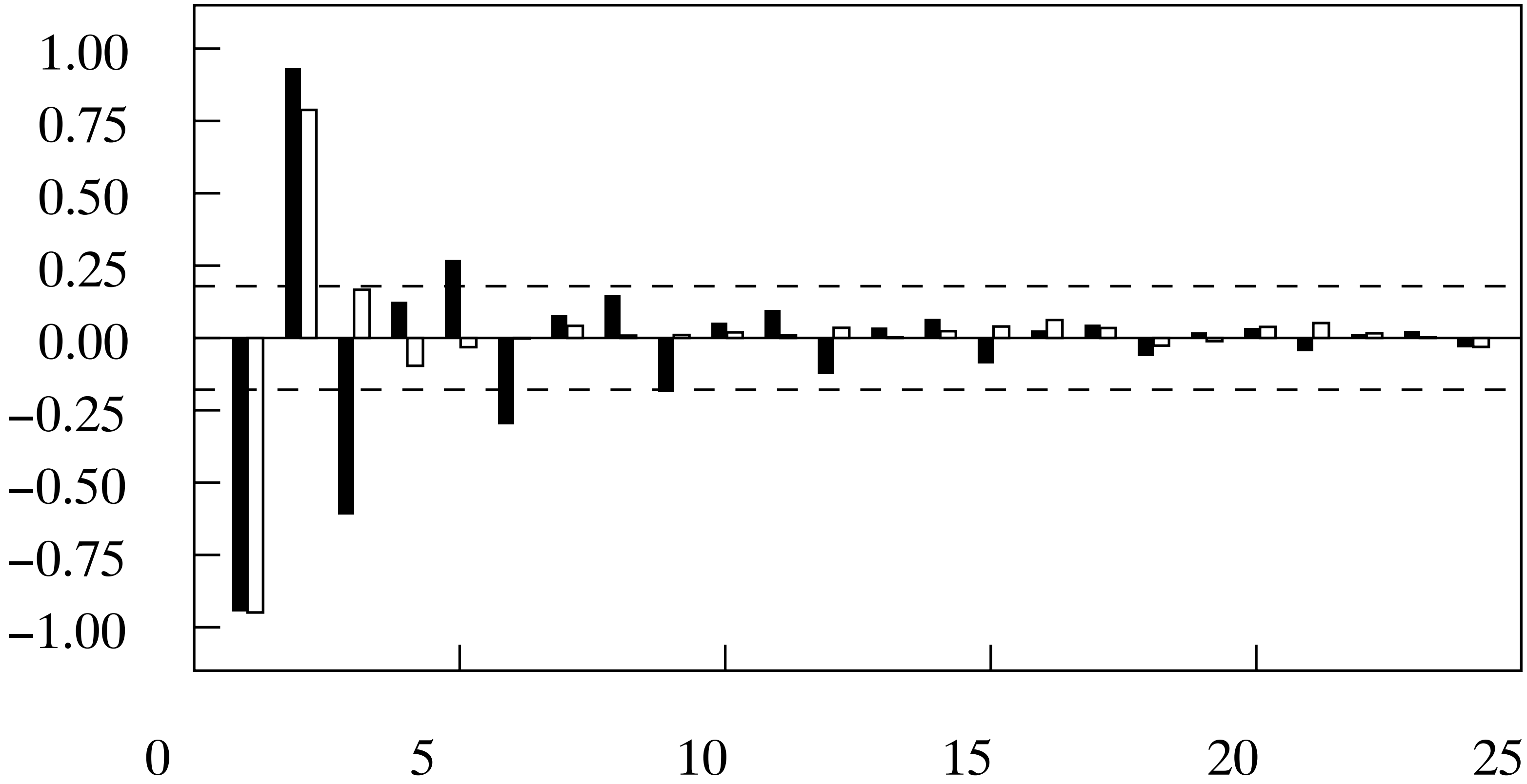

Figure 18. The theoretical partial autocorrelation function (PACF) of the ARMA(2, 2) process $(1 - 1.69L + 0.81L^2)y(t) = (1 + 0.90L + 0.81L^2)\varepsilon(t)$ (the solid bars) together with its empirical counterpart, calculated from a simulated series of 120 observations..

FORECASTING WITH ARMA MODELS

The Coefficients of the Moving-Average Expansion

The ARMA model $\alpha(L)y(t) = \mu(L)\varepsilon(t)$ can be cast in the form of $y(t) = \{\mu(L)/\alpha(L)\}\varepsilon(t) = \psi(L)\varepsilon(t)$ where

$$\psi(L) = \{\psi_0 + \psi_1 L + \psi_2 L^2 + \cdots\}$$

is from the expansion of the rational function.

The method of finding the coefficients of the series expansion can be illustrated by the second-order case:

$$\frac{\mu_0 + \mu_1 z}{\alpha_1 + \alpha_1 z + \alpha_2 z^2} = \Big\{\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots\Big\}.$$

We rewrite this equation as

$$\mu_0 + \mu_1 z = \Big\{\alpha_1 + \alpha_1 z + \alpha_2 z^2\Big\}\Big\{\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots\Big\}.$$

The following table assists us in multipling togther the two polyomials:

	ψ_0	$\psi_1 z$	$\psi_2 z^2$	$\cdots$
α_0	$\alpha_0\psi_0$	$\alpha_0\psi_1 z$	$\alpha_0\psi_2 z^2$	$\cdots$
$\alpha_1 z$	$\alpha_1\psi_0 z$	$\alpha_1\psi_1 z^2$	$\alpha_1\psi_2 z^3$	$\cdots$
$\alpha_2 z^2$	$\alpha_2\psi_0 z^2$	$\alpha_2\psi_1 z^3$	$\alpha_2\psi_2 z^4$	$\cdots$

Performing the multiplication on the RHS of the equation, and by equating the coefficients of the same powers of z on the two sides, we find that

$$\begin{aligned}
\mu_0 &= \alpha_0\psi_0, & \psi_0 &= \mu_0/\alpha_0,\\
\mu_1 &= \alpha_0\psi_1 + \alpha_1\psi_0, & \psi_1 &= (\mu_1 - \alpha_1\psi_0)/\alpha_0,\\
0 &= \alpha_0\psi_2 + \alpha_1\psi_1 + \alpha_2\psi_0, & \psi_2 &= -(\alpha_1\psi_1 + \alpha_2\psi_0)/\alpha_0,\\
&\vdots & &\vdots\\
0 &= \alpha_0\psi_n + \alpha_1\psi_{n-1} + \alpha_2\psi_{n-2}, & \psi_n &= -(\alpha_1\psi_{n-1} + \alpha_2\psi_{n-2})/\alpha_0.
\end{aligned}$$

The optimal (minimum mean-square error) forecast of y_{t+h} is the conditional expectation of y_{t+h} given the information set $\mathcal{I}_t$ comprising the values of $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ or equally the values of $\{y_t, y_{t-1}, y_{t-2}, \ldots\}$. On taking expectations $y(t)$ and $\varepsilon(t)$ conditonal on $\mathcal{I}_t$, we find that

$$
\begin{aligned}
E(y_{t+k}|\mathcal{I}_t) &= \hat{y}_{t+k} \quad \text{if} \quad k > 0, \\
E(y_{t-j}|\mathcal{I}_t) &= y_{t-j} \quad \text{if} \quad j \geq 0, \\
E(\varepsilon_{t+k}|\mathcal{I}_t) &= 0 \quad \text{if} \quad k > 0, \\
E(\varepsilon_{t-j}|\mathcal{I}_t) &= \varepsilon_{t-j} = y_{t-j} - \hat{y}_{t-j} \quad \text{if} \quad j \geq 0.
\end{aligned} \tag{21}
$$

In this notation, the forecast h periods ahead is

$$
\begin{aligned}
E(y_{t+h}|\mathcal{I}_t) &= \sum_{k=1}^{h} \psi_{h-k} E(\varepsilon_{t+k}|\mathcal{I}_t) + \sum_{j=0}^{\infty} \psi_{h+j} E(\varepsilon_{t-j}|\mathcal{I}_t) \\
&= \sum_{j=0}^{\infty} \psi_{h+j} \varepsilon_{t-j}.
\end{aligned} \tag{22}
$$

In practice, the forecasts are generated recursively via the equation

$$
\begin{aligned}
y(t) = -\{\alpha_1 y(t-1) + \alpha_2 y(t-2) + \cdots + \alpha_p y(t-p)\}\\
+ \mu_0\varepsilon(t) + \mu_1\varepsilon(t-1) + \cdots + \mu_q\varepsilon(t-q).
\end{aligned} \tag{23}
$$

By taking the conditional expectation of this function, we get

$$
\begin{aligned}
\hat{y}_{t+h} = -\{\alpha_1\hat{y}_{t+h-1} + \cdots + \alpha_p y_{t+h-p}\}\\
+ \mu_h\varepsilon_t + \cdots + \mu_q\varepsilon_{t+h-q} \quad \text{when} \quad 0 < h \leq p, q,
\end{aligned} \tag{24}
$$

$$
\hat{y}_{t+h} = -\{\alpha_1\hat{y}_{t+h-1} + \cdots + \alpha_p y_{t+h-p}\} \quad \text{if} \quad q < m \leq p, \tag{25}
$$

$$
\begin{aligned}
\hat{y}_{t+h} = -\{\alpha_1\hat{y}_{t+h-1} + \cdots + \alpha_p\hat{y}_{t+h-p}\}\\
+ \mu_h\varepsilon_t + \cdots + \mu_q\varepsilon_{t+h-q} \quad \text{if} \quad p < h \leq q,
\end{aligned} \tag{26}
$$

and

$$
\hat{y}_{t+h} = -\{\alpha_1\hat{y}_{t+h-1} + \cdots + \alpha_p\hat{y}_{t+h-p}\} \quad \text{when} \quad p, q < h. \tag{27}
$$

Equation (27) ashows that, whn $h > p, q$, the forecasting function becomes a pth-order homogeneous difference equation in y. The p values of $y(t)$ from $t = r = \max(p, q)$ to $t = r - p + 1$ serve as the starting values for the equation.

The behaviour of the forecast function beyond the reach of the starting values is determind the roots of the autoregressive operator $\alpha(L) = 0$

If all of the roots of $\alpha(z) = 0$ are less than unity, then $\hat{y}_{t+h}$ will converge to zero as h increases.

If one of the roots is unity, then the forecast function will converge to a nonzero consant.

If the are two unit roots, then the forecast function will converg to a linear trend.

In general, if d of the roots are unity, then the general solution will comprise a polynomial in t of order $d - 1$.

The forecasts can be updated easily once the coefficients in the expansion of $\psi(L) = \mu(L)/\alpha(L)$ have been obtained. Consider

$$(28)\qquad \begin{aligned} \hat{y}_{t+h|t+1} &= \{\psi_{h-1}\varepsilon_{t+1} + \psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \cdots\} \quad \text{and} \\ \hat{y}_{t+h|t} &= \{\psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \psi_{h+2}\varepsilon_{t-2} + \cdots\}. \end{aligned}$$

The first of these is the forecast for $h-1$ periods ahead made at time $t+1$ whilst the second is the forecast for h periods ahead made at time t. It can be seen that

$$(29)\qquad \hat{y}_{t+h|t+1} = \hat{y}_{t+h|t} + \psi_{h-1}\varepsilon_{t+1},$$

where $\varepsilon_{t+1} = y_{t+1} - \hat{y}_{t+1}$ is the current disturbance at time $t+1$. The later is also the prediction error of the one-step-ahead forecast made at time t.