

APPENDIX 4

The Identification of ARIMA Models

As we have established in a previous lecture, there is a one-to-one correspondence between the parameters of an ARMA(p, q) model, including the variance of the disturbance, and the leading $p + q + 1$ elements of the autocovariance function. Given the true autocovariances of a process, we might be able to discern the orders p and q of its autoregressive and moving-average operators and, given these orders, we should then be able to deduce the values of the parameters.

There are two other functions, prominent in time-series analysis, from which it is possible to recover the parameters of an ARMA process. These are the partial autocorrelation function and the spectral density function. The appearance of each of these functions gives an indication of the nature of the underlying process to which they belong; and, in theory, the business of identifying the model and of recovering its parameters can be conducted on the basis of any of them. In practice, the process is assisted by taking account of all three functions.

The empirical versions of the three functions which are used in a model-building exercise may differ considerably from their theoretical counterparts. Even when the data are truly generated by an ARMA process, the sampling errors which affect the empirical functions can lead one to identify the wrong model. This hazard is revealed by sampling experiments. When the data come from the real world, the notion that there is an underlying ARMA process is a fiction, and the business of model identification becomes more doubtful. Then there may be no such thing as the correct model; and the choice amongst alternative models must be made partly with a view their intended uses.

The Autocorrelation Functions

The techniques of model identification which are most commonly used were propounded originally by Box and Jenkins (1972). Their basic tools were the sample autocorrelation function and the partial autocorrelation function. We shall describe these functions and their use separately from the spectral density function which ought, perhaps, to be used more often in selecting models. The fact that spectral density function is often overlooked is probably due to

an unfamiliarity with frequency-domain analysis on the part of many model builders.

Autocorrelation function (ACF). Given a sample y_0, y_1, \dots, y_{T-1} of T observations, we define the sample autocorrelation function to be the sequence of values

$$(1) \quad r_\tau = c_\tau / c_0, \quad \tau = 0, 1, \dots, T - 1,$$

wherein

$$(2) \quad c_\tau = \frac{1}{T} \sum_{t=\tau}^{T-1} (y_t - \bar{y})(y_{t-\tau} - \bar{y})$$

is the empirical autocovariance at lag τ and c_0 is the sample variance. One should note that, as the value of the lag increases, the number of observations comprised in the empirical autocovariance diminishes until the final element $c_{T-1} = T^{-1}(y_0 - \bar{y})(y_{T-1} - \bar{y})$ is reached which comprises only the first and last mean-adjusted observations.

In plotting the sequence $\{r_\tau\}$, we shall omit the value of r_0 which is invariably unity. Moreover, in interpreting the plot, one should be wary of giving too much credence to the empirical autocorrelations at lag values which are significantly high in relation to the size of the sample.

Partial autocorrelation function (PACF). The sample partial autocorrelation p_τ at lag τ is simply 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 $y_1, y_2, \dots, y_{t-\tau+1}$. The partial autocorrelation measures the dependence between y_t and $y_{t-\tau}$ after the effect of the intervening values has been removed.

The sample partial autocorrelation p_τ is virtually the same quantity as the estimated coefficient of lag τ obtained by fitting an autoregressive model of order τ to the data. Indeed, the difference between the two quantities vanishes as the sample size increases. The Durbin–Levinson algorithm provides an efficient way of computing the sequence $\{p_\tau\}$ of partial autocorrelations from the sequence of $\{c_\tau\}$ of autocovariances. It can be seen, in view of this algorithm, that the information in $\{c_\tau\}$ is equivalent to the information contained jointly in $\{p_\tau\}$ and c_0 . Therefore the sample autocorrelation function $\{r_t\}$ and the sample partial autocorrelation function $\{p_t\}$ are equivalent in terms of their information content.

The Methodology of Box and Jenkins

The model-building methodology of Box and Jenkins, relies heavily upon the two functions $\{r_t\}$ and $\{p_t\}$ defined above. It involves a cycle comprising the three stages of model selection, model estimation and model checking. In view of the difficulties of selecting an appropriate model, it is envisaged that the cycle might have to be repeated several times and that, at the end, there might be more than one model of the same series.

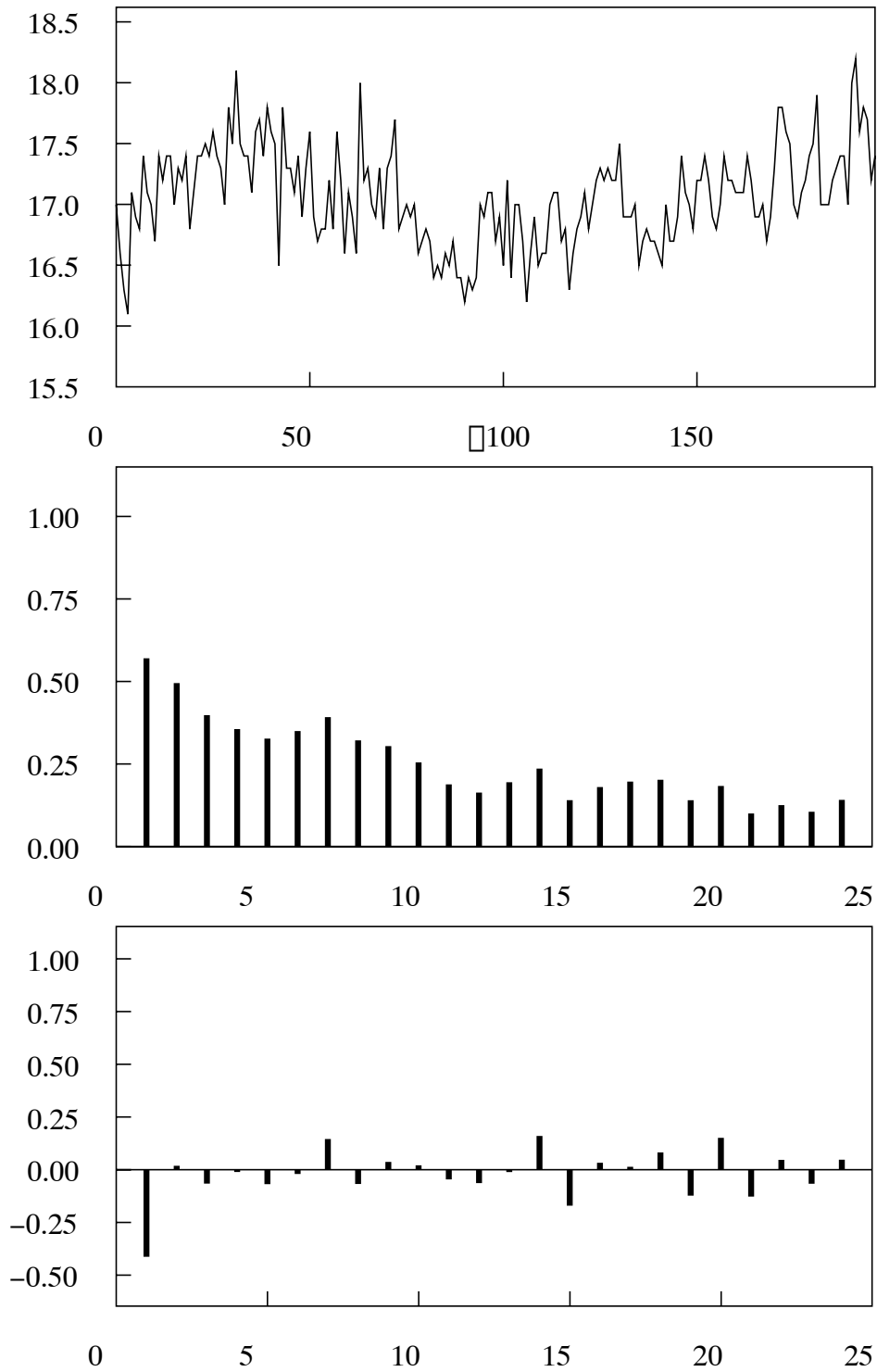


Figure 1. The concentration readings from a chemical process with the autocorrelation function and the autocorrelation function of the differences.

Reduction to stationarity. The first step, which is taken before embarking on the cycle, is to examine the time plot of the data and to judge whether or not it could be the outcome of a stationary process. If a trend is evident in the data, then it must be removed. A variety of techniques of trend removal, which include the fitting of parametric curves and of spline functions, have been discussed in previous lectures. When such a function is fitted, it is to the sequence of residuals that the ARMA model is applied.

However, Box and Jenkins were inclined to believe that many empirical series can be modelled adequately by supposing that some suitable difference of the process is stationary. Thus the process generating the observed series $y(t)$ might be modelled by the ARIMA(p, d, q) equation

$$(3) \quad \alpha(L)\nabla^d y(t) = \mu(L)\varepsilon(t),$$

wherein $\nabla^d = (I - L)^d$ is the d th power of the difference operator. In that case, the differenced series $z(t) = \nabla^d y(t)$ will be described by a stationary ARMA(p, q) model. The inverse operator ∇^{-1} is the summing or integrating operator, which accounts for the fact that the model depicted by equation (3) is described an autoregressive integrated moving-average model.

To determine whether stationarity has been achieved, either by trend removal or by differencing, one may examine the autocorrelation sequence of the residual or processed series. The sequence corresponding to a stationary process should converge quite rapidly to zero as the value of the lag increases. An empirical autocorrelation function which exhibits a smooth pattern of significant values at high lags indicates a nonstationary series.

An example is provided by Figure 1 where a comparison is made between the autocorrelation function of the original series and that of its differences. Although the original series does not appear to embody a systematic trend, it does drift in a haphazard manner which suggests a random walk; and it is appropriate to apply the difference operator.

Once the degree of differencing has been determined, the autoregressive and moving-average orders are selected by examining the sample autocorrelations and sample partial autocorrelations. 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 a pure moving-average process of order q has $\rho_\tau = 0$ for all $\tau > q$. The corresponding partial autocorrelation function $\{\pi_\tau\}$ is liable to decay towards zero gradually. To judge whether the corresponding sample autocorrelation function $\{r_\tau\}$ shows evidence of a truncation, we need some scale by which to judge the significance of the values of its elements.

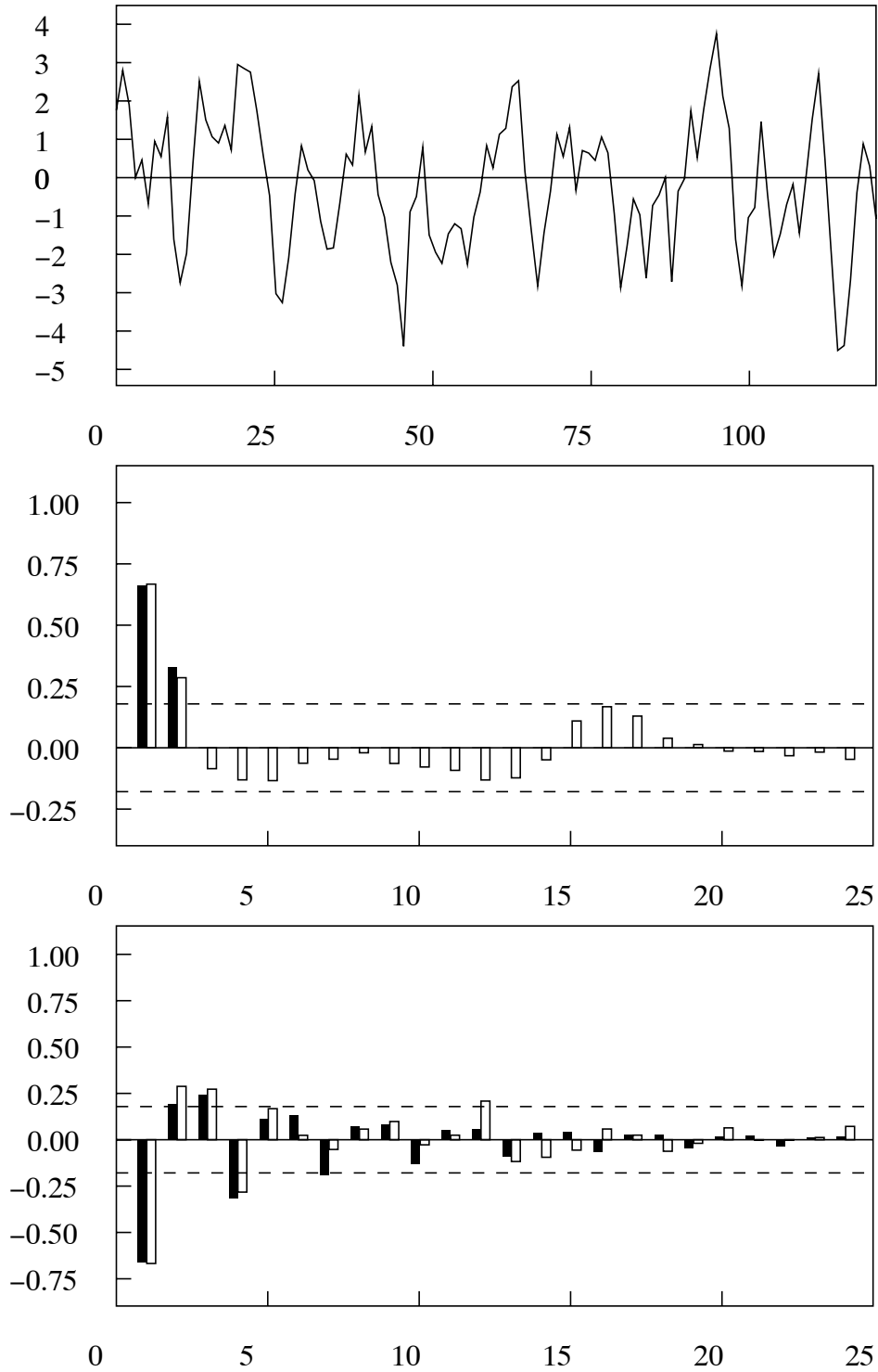


Figure 2. The graph of 120 observations on a simulated series generated by the MA(2) process $y(t) = (1 + 0.90L + 0.81L^2)\varepsilon(t)$ together with the theoretical and empirical ACF's (middle) and the theoretical and empirical PACF's (bottom). The theoretical values correspond to the solid bars.

As a guide to determining whether the parent autocorrelations are in fact 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

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

The result is also given by Fuller [1976, p. 237]. A simpler 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.

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

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

In general, the sequence generated by this equation will represent 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. One can expect the empirical autocovariance function of a pure AR process to be of the same nature as its theoretical parent.

It is the partial autocorrelation function which serves most clearly to identify a pure AR process. The theoretical partial autocorrelations function $\{\pi_\tau\}$ of a $\text{AR}(p)$ process has $\pi_\tau = 0$ for all $\tau > p$. Likewise, all elements of the sample partial autocorrelation function are expected to be close to zero for lags greater than p , which corresponds to the fact that they are simply estimates of zero-valued parameters. The significance of the values of the partial autocorrelations is judged by the fact that, for a p th order process, their standard deviations for all lags greater than p are approximated by $1/\sqrt{T}$. Thus the bounds of $\pm 1.96/\sqrt{T}$ are also plotted on the graph of the partial autocorrelation function.

Mixed processes. In the case of a mixed $\text{ARMA}(p, q)$ process, neither the theoretical autocorrelation function nor the theoretical partial autocorrelation function have any abrupt cutoffs. Indeed, there is little that can be inferred from either of these functions or from their empirical counterparts beyond the fact that neither a pure MA model nor a pure AR model would be inappropriate. On its own, the autocovariance function of an $\text{ARMA}(p, q)$ process is not easily distinguished from that of a pure AR process. In particular, its elements γ_τ satisfy the same difference equation as that of a pure AR model for all values of $\tau > \max(p, q)$.

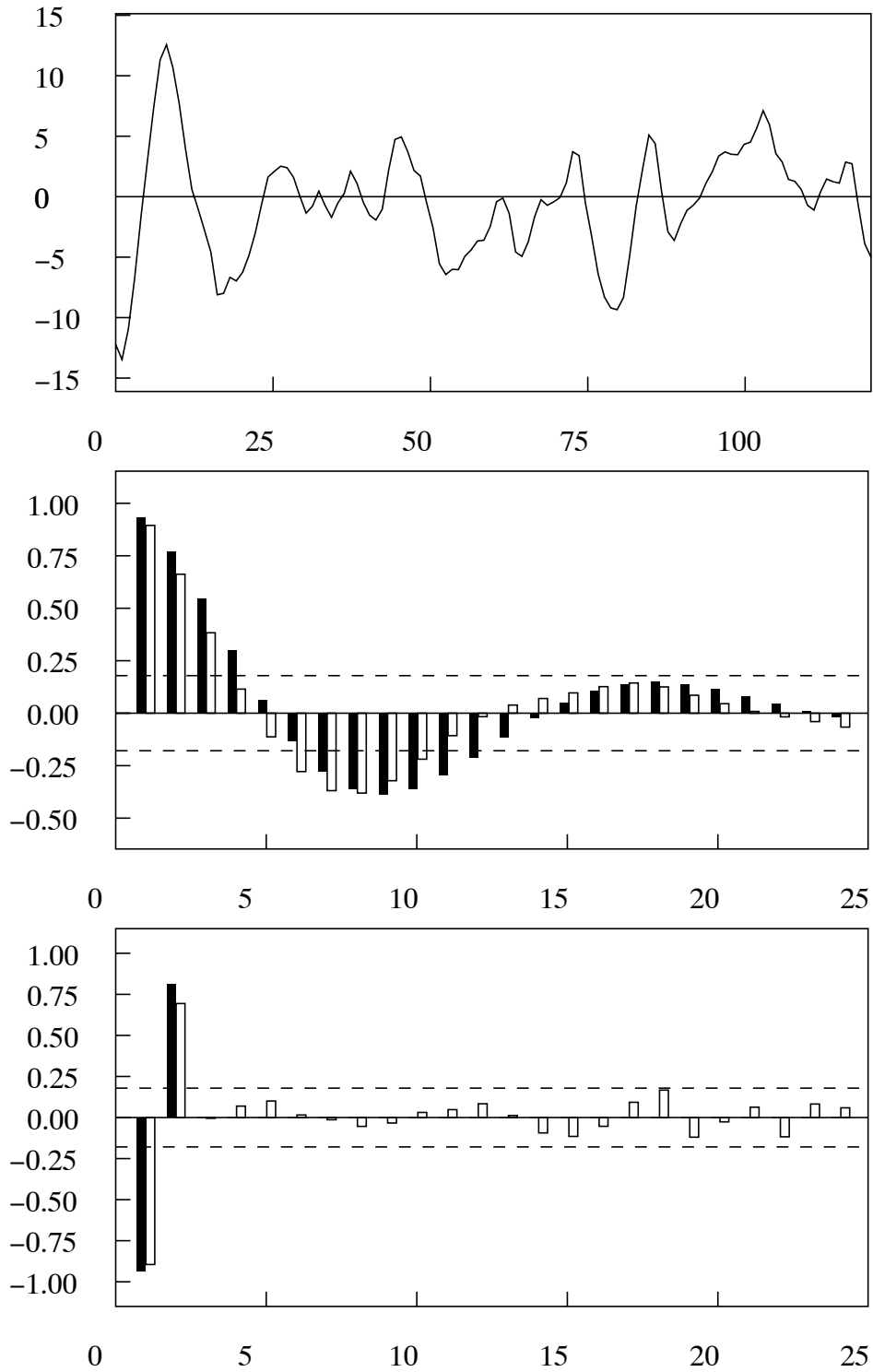


Figure 3. The graph of 120 observations on a simulated series generated by the AR(2) process $(1 - 1.69L + 0.81L^2)y(t) = \varepsilon(t)$ together with the theoretical and empirical ACF's (middle) and the theoretical and empirical PACF's (bottom). The theoretical values correspond to the solid bars.

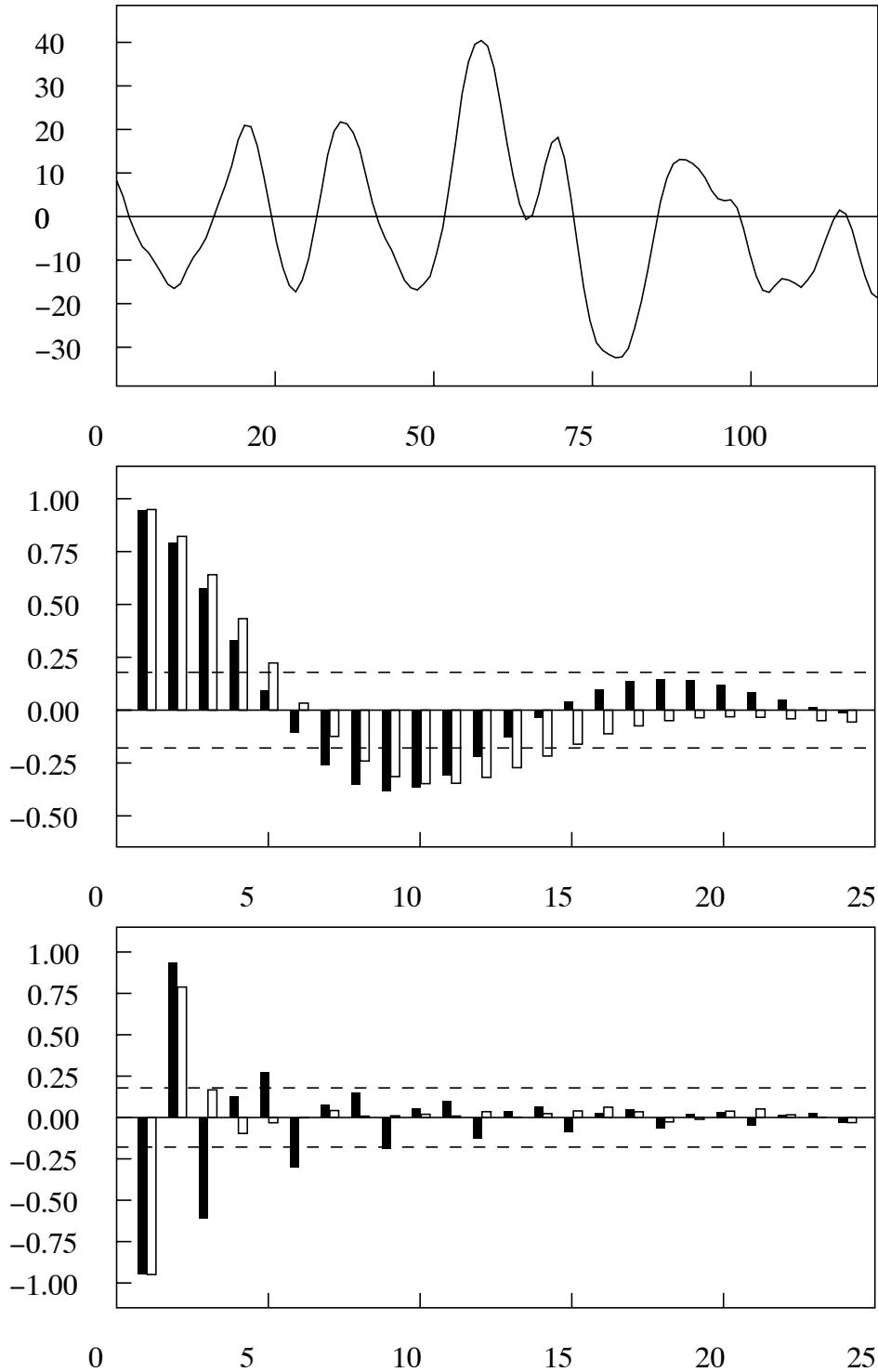


Figure 4. The graph of 120 observations on a simulated series generated by the ARMA(2, 2) process $(1 - 1.69L + 0.81L^2)y(t) = (1 + 0.90L + 0.81L^2)\varepsilon(t)$ together with the theoretical and empirical ACF's (middle) and the theoretical and empirical PACF's (bottom). The theoretical values correspond to the solid bars.

There is good reason to regard mixed models as more appropriate in practice than pure models of either variety. For a start, there is the fact that a rational transfer function is far more effective in approximating an arbitrary impulse response than is an autoregressive transfer function, whose parameters are confined to the denominator, or a moving-average transfer function, which has its parameters in the numerator. Indeed, it might be appropriate, sometimes, to approximate a pure process of a high order by a more parsimonious mixed model.

Mixed models are also favoured by the fact that the sum of any two mutually independent autoregressive process gives rise to an ARMA process. Let $y(t)$ and $z(t)$ be autoregressive processes of orders p and r respectively which are described by the equations $\alpha(L)y(t) = \varepsilon(t)$ and $\rho(L)z(t) = \eta(t)$, wherein $\varepsilon(t)$ and $\eta(t)$ are mutually independent white-noise processes. Then their sum will be

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

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

In economics, where the data series are highly aggregated, mixed models would seem to be called for often. In the context of electrical and mechanical engineering, there may be some justification for pure AR models. Here there is often abundant data, sufficient to sustain the estimation of pure autoregressive models of high order. Therefore the principle of parametric parsimony is less persuasive than it might be in an econometric context. However, pure AR models perform poorly whenever the data is affected by errors of observation; and, in this respect, a mixed model is liable to be more robust. One can understand this feature of mixed models by recognising that the sum of a pure $AR(p)$ process and a white-noise process is an $ARMA(p, p)$ process.