The loglikelihood function for the first p observations is the logarithm of the joint density of the vector $\boldsymbol{y}^p$, which consists of the first p observations on y_t. If we let $\omega^2\boldsymbol{\Delta}_p$ denote the $p \times p$ covariance matrix of the first p u_t's and let $\boldsymbol{x}^p(\boldsymbol{\beta})$ denote the first p observations on $x_t(\boldsymbol{\beta})$, it will be

$$\begin{aligned}\ell^p(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \omega) = & -\tfrac{p}{2}\log(2\pi) - p\log(\omega) + \tfrac{1}{2}\log|\boldsymbol{\Delta}_p^{-1}| \\ & - \frac{1}{2\omega^2}\big(\boldsymbol{y}^p - \boldsymbol{x}^p(\boldsymbol{\beta})\big)^{\top}\boldsymbol{\Delta}_p^{-1}\big(\boldsymbol{y}^p - \boldsymbol{x}^p(\boldsymbol{\beta})\big).\end{aligned} \tag{10.56}$$

If $p = 1$, $|\boldsymbol{\Delta}_p^{-1}| = \boldsymbol{\Delta}_p^{-1} = 1 - \rho^2$. Thus (10.50) is seen to be a special case of (10.56).

The full loglikelihood function is the sum of (10.55) and (10.56). As in the AR(1) case, the presence of the Jacobian term $\frac{1}{2}\log|\boldsymbol{\Delta}_p^{-1}|$ ensures that this function will have at least one maximum within the stationarity region. However, it also makes evaluating and maximizing the function a good deal more difficult. Some authors (e.g., Box and Jenkins (1976)) have therefore suggested ignoring it and maximizing the rest of the loglikelihood function. Other references on the estimation of models with AR(p) errors include Ansley (1979), Kendall, Stuart, and Ord (1983), and Granger and Newbold (1986). Beach and MacKinnon (1978b) discuss the AR(2) case in some detail.

10.7 Moving Average and ARMA Processes

Autoregressive processes are not the only way to model stationary time series. The other basic type of stochastic process is the **moving average**, or **MA**, process. The simplest moving average process is the **first-order moving average**, or **MA(1)**, process

$$u_t = \varepsilon_t + \alpha_1\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathrm{IID}(0, \omega^2), \tag{10.57}$$

in which the error u_t is literally a moving average of two successive innovations, ε_t and ε_{t-1}. Thus ε_t affects both u_t and u_{t+1} but does not affect u_{t+j} for $j > 1$. The more general MA(q) process may be written either as

$$u_t = \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + \cdots + \alpha_q\varepsilon_{t-q}, \;\; \varepsilon_t \sim \mathrm{IID}(0, \omega^2)$$

or, using lag-operator notation, as

$$u_t = \big(1 + \alpha_1 L + \cdots + \alpha_q L^q\big)\varepsilon_t \equiv B(L, \boldsymbol{\alpha})\varepsilon_t, \;\; \varepsilon_t \sim \mathrm{IID}(0, \omega^2), \tag{10.58}$$

where $\boldsymbol{\alpha} \equiv [\alpha_1 \;\vdots\; \alpha_2 \;\vdots\; \cdots \;\vdots\; \alpha_q]$.

Finite-order MA processes are necessarily stationary, since each u_t is a weighted sum of a finite number of innovations ε_t, $\varepsilon_{t-1} \cdots$. Thus we do not have to impose stationarity conditions. We do, however, have to impose an