paragraph. When these two equations are solved for v_0 and v_1, they yield

$$v_0 = \sigma_\varepsilon^2 \, \frac{1 + 2\rho_1\alpha_1 + \alpha_1^2}{1 - \rho_1^2}, \;\text{ and }\; v_1 = \sigma_\varepsilon^2 \, \frac{(\rho_1 + \alpha_1)(1 + \rho_1\alpha_1)}{1 - \rho_1^2}. \tag{13.24}$$

Finally, multiplying equation (13.23) by u_{t-i} for $i > 1$ and taking expectations gives $v_i = \rho_1 v_{i-1}$, from which we conclude that

$$v_i = \sigma_\varepsilon^2 \rho_1^{i-1} \frac{(\rho_1 + \alpha_1)(1 + \rho_1\alpha_1)}{1 - \rho_1^2}. \tag{13.25}$$

Equation (13.25) provides all the autocovariances of an ARMA$(1, 1)$ process. Using it and the first of equations (13.24), we can derive the autocorrelations.

Autocorrelation Functions

As we have seen, the autocorrelation between u_t and u_{t-j} can be calculated theoretically for any known stationary ARMA process. The **autocorrelation function**, or **ACF**, expresses the autocorrelation as a function of the lag j for $j = 1, 2 \ldots.$ If we have a sample y_t, $t = 1, \ldots, n$, from an ARMA process of possibly unknown order, then the j^{th} order autocorrelation $\rho(j)$ can be estimated by using the formula

$$\hat\rho(j) = \frac{\widehat{\text{Cov}}(y_t, y_{t-j})}{\widehat{\text{Var}}(y_t)}, \tag{13.26}$$

where

$$\widehat{\text{Cov}}(y_t, y_{t-j}) = \frac{1}{n-1} \sum_{t=j+1}^{n} (y_t - \bar y)(y_{t-j} - \bar y), \tag{13.27}$$

and

$$\widehat{\text{Var}}(y_t) = \frac{1}{n-1} \sum_{t=1}^{n} (y_t - \bar y)^2. \tag{13.28}$$

In equations (13.27) and (13.28), $\bar y$ is the mean of the y_t. Of course, (13.28) is just the special case of (13.27) in which $j = 0$. It may seem odd to divide by $n - 1$ rather than by $n - j - 1$ in (13.27). However, if we did not use the same denominator for every j, the estimated autocorrelation matrix would not necessarily be positive definite. Because the denominator is the same, the factors of $1/(n - 1)$ cancel in the formula (13.26).

The **empirical ACF**, or **sample ACF**, expresses the $\hat\rho(j)$, defined in equation (13.26), as a function of the lag j. Graphing the sample ACF provides a convenient way to see what the pattern of serial dependence in any observed time series looks like, and it may help to suggest what sort of stochastic process would provide a good way to model the data. For example, if the data were generated by an MA(1) process, we would expect that $\hat\rho(1)$ would be an estimate of $\alpha_1/(1+\alpha_1^2)$ and all the other $\hat{\rho}(j)$ would be approximately equal to zero. If the data were generated by an AR(1) process with $\rho_1 > 0$, we would expect that $\hat{\rho}(1)$ would be an estimate of ρ_1 and would be relatively large, the next few $\hat{\rho}(j)$ would be progressively smaller, and the ones for large j would be approximately equal to zero. A graph of the sample ACF is sometimes called a **correlogram**; see Exercise 13.15.

The **partial autocorrelation function**, or **PACF**, is another way to characterize the relationship between y_t and its lagged values. The partial autocorrelation coefficient of order j is defined as the plim of the least squares estimator of the coefficient $\rho_j^{(j)}$ in the linear regression

$$y_t = \gamma^{(j)} + \rho_1^{(j)} y_{t-1} + \ldots + \rho_j^{(j)} y_{t-j} + \varepsilon_t, \tag{13.29}$$

or, equivalently, in the minimization problem

$$\min_{\gamma^{(j)},\, \rho_i^{(j)}} \mathrm{E}\Big(y_t - \gamma^{(j)} - \sum_{i=1}^{j} \rho_i^{(j)} y_{t-i}\Big)^2. \tag{13.30}$$

The superscript "(j)" appears on all the coefficients in regression (13.29) to make it plain that all the coefficients, not just the last one, are functions of j, the number of lags. We can calculate the **empirical PACF**, or **sample PACF**, up to order J by running regression (13.29) for $j = 1, \ldots, J$ and retaining only the estimate $\hat{\rho}_j^{(j)}$ for each j. Just as a graph of the sample ACF may help to suggest what sort of stochastic process would provide a good way to model the data, so a graph of the sample PACF, interpreted properly, may do the same. For example, if the data were generated by an AR(2) process, we would expect the first two partial autocorrelations to be relatively large, and all the remaining ones to be insignificantly different from zero.

13.3 Estimating AR, MA, and ARMA Models

All of the time-series models that we have discussed so far are special cases of an ARMA(p, q) model with a constant term, which can be written as

$$y_t = \gamma + \sum_{i=1}^{p} \rho_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}, \tag{13.31}$$

where the ε_t are assumed to be white noise. Not counting the variance of the innovations, there are $p + q + 1$ parameters to estimate in the model (13.31): the ρ_i, for $i = 1, \ldots, p$, the α_j, for $j = 1, \ldots, q$, and γ. Recall that γ is not the unconditional expectation of y_t unless all of the ρ_i are zero.

For our present purposes, it is perfectly convenient to work with models that allow y_t to depend on exogenous explanatory variables and are therefore even