A possible solution arises if we write ε_{t-1} in this expression as a function of observed Y_ts. This is possible only if the MA polynomial is *invertible*. In this case we can use that

$$\varepsilon_{t-1} = \sum_{j=0}^{\infty} (-\alpha)^j (Y_{t-j-1} - \mu)$$

(see above) and write

$$S(\alpha, \mu) = \sum_{t=2}^{T} \left(Y_t - \mu - \alpha \sum_{j=0}^{\infty} (-\alpha)^j (Y_{t-j-1} - \mu) \right)^2 .$$

In practice, Y_t is not observed for $t = 0, -1, \ldots$, so we have to cut off the infinite sum in this expression to obtain an approximate sum of squares

$$\tilde{S}(\alpha, \mu) = \sum_{t=2}^{T} \left(Y_t - \mu - \alpha \sum_{j=0}^{t-2} (-\alpha)^j (Y_{t-j-1} - \mu) \right)^2 . \qquad (8.62)$$

Because, asymptotically, if T goes to infinity the difference between $S(\alpha, \mu)$ and $\tilde{S}(\alpha, \mu)$ disappears minimizing (8.62) with respect to α and μ gives consistent estimators $\hat{\alpha}$ and $\hat{\mu}$. Unfortunately, (8.62) is a high order polynomial in α and thus has very many local minima. Therefore, numerically minimizing (8.62) is complicated. However, as we know that $-1 < \alpha < 1$, a grid search (e.g. $-0.99, -0.98, -0.97, \ldots, 0.98, 0.99$) can be performed. The resulting nonlinear least squares estimators for α and μ are consistent and asymptotically normal.

8.6.2 Maximum Likelihood

An alternative estimator for ARMA models is provided by maximum likelihood. This requires that an assumption is made about the distribution of ε_t, most commonly normality. Although the normality assumption is strong, the ML estimators are very often consistent even in cases where ε_t is not normal. Conditional upon an initial value the loglikelihood function can be written as

$$\log L(\alpha, \theta, \mu, \sigma^2) = -\frac{T-1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{t=2}^{T} \varepsilon_t^2/\sigma^2,$$

where ε_t is a function of the coefficients α, θ and μ, y_t and its history. For an $AR(1)$ model it is $\varepsilon_t = y_t - \theta y_{t-1}$, where $y_t = Y_t - \mu$, and for the $MA(1)$ model we have

$$\varepsilon_t = y_t - \alpha \sum_{j=0}^{t-2} (-\alpha)^j y_{t-j-1} = \sum_{j=0}^{t-1} (-\alpha)^j y_{t-j}.$$

Both of the implied loglikelihood functions are conditional upon an initial value. For the $AR(1)$ case y_1 is treated as given, while for the $MA(1)$ case the initial condition is $\varepsilon_0 = 0$. The resulting estimators are therefore referred to as **conditional maximum likelihood** estimators. The conditional ML estimators for α, θ and μ are easily seen to be identical to the least squares estimators.

The *exact* maximum likelihood estimator combines the conditional likelihood with the likelihood from the initial observations. In the $AR(1)$ case, for example, the following term is added to the loglikelihood:

$$-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log[\sigma^2/(1-\theta^2)] - \frac{1}{2}\frac{y_1^2}{\sigma^2/(1-\theta^2)},$$

which follows from the fact that the marginal density of y_1 is normal with mean zero and variance $\sigma^2/(1-\theta^2)$. For a moving average process the exact likelihood function is somewhat more complex. If T is large the way we treat the initial values has negligible impact, so that the conditional and exact maximum likelihood estimators are asymptotically equivalent in cases where the AR and MA polynomials are invertible. More details can be found in Hamilton (1994, Chapter 5).

It will be clear from the results above that estimating autoregressive models is simpler than estimating moving average models. Estimating ARMA models, which combine an autoregressive part with a moving average part, closely follows the lines of ML estimation of the MA parameters. As any (invertible) ARMA model can be approximated by an autoregressive model of infinite order, it has become more and more common practice to use autoregressive specifications instead of MA or ARMA ones, and allowing for a sufficient number of lags. Particularly if the number of observations is not too small this approach may work pretty well in practice. Of course, an MA representation of the same process may be more parsimonious. Another advantage of autoregressive models is that they are easily generalized to multivariate time series, where one wants to model a set of economic variables jointly. This leads to so-called **vector autoregressive models** (VAR models), which are discussed in the next chapter.

8.7 Choosing a Model

Most of the time there are no economic reasons to choose a particular specification of the model. Consequently, to a large extent the data will determine which time series model is appropriate. Before estimating any model, it is common to estimate autocorrelation and partial autocorrelation coefficients directly from the data. Often this gives some idea about which model might be appropriate. After one or more models are estimated, their quality can be judged by checking whether the residuals are more or less white noise, and by comparing them with alternative specifications. These comparisons can be based on statistical significance tests or the use of particular model selection criteria.

8.7.1 The Autocorrelation Function

The autocorrelation function (ACF) describes the correlation between Y_t and its lag Y_{t-k} as a function of k. Recall that the k-th order autocorrelation coefficient is defined as

$$\rho_k = \frac{\text{cov}\{Y_t, Y_{t-k}\}}{V\{Y_t\}} = \frac{\gamma_k}{\gamma_0},$$

noting that $\text{cov}\{Y_t, Y_{t-k}\} = E\{y_t y_{t-k}\}$.

Table 8.3 Unit root tests for log exchange rate Italy–France

Statistic	Without trend	With trend
DF	−0.328	−1.900
ADF(1)	−0.361	−1.884
ADF(2)	−0.160	−1.925
ADF(3)	−0.291	−2.012
ADF(4)	−0.366	−2.026
ADF(5)	−0.463	−2.032
ADF(6)	−0.643	−2.262

ADF tests up to lag 6. The results here are quite clear. In none of the cases can we reject the null hypothesis of a unit root.

If purchasing power parity between France and Italy holds in the long run, one can expect that short-run deviations, $s_t - (p_t - p_t^*)$, corresponding to the real exchange rate, are limited and do not wander widely. In other words, one can expect $s_t - (p_t - p_t^*)$ to be stationary. A test for PPP can thus be based on the analysis of the log real exchange rate $rs_t \equiv s_t - (p_t - p_t^*)$. The series is plotted in Figure 8.6, while the results for the augmented Dickey–Fuller tests for this variable are given in Table 8.4.

The results show that the null hypothesis of a unit root in rs_t (corresponding to non-stationarity) cannot be rejected. Consequently, there is no evidence for PPP to hold in this form. One reason why we may not be able to reject the null hypothesis is simply that our sample contains insufficient information, that is: our sample is too short and standard errors are simply too high to reject the unit root hypothesis. This is a problem often found in tests for purchasing power parity. A critical survey of this literature can be found in Froot and Rogoff (1996). In the next chapter, we shall also analyse whether some weaker form of PPP holds.

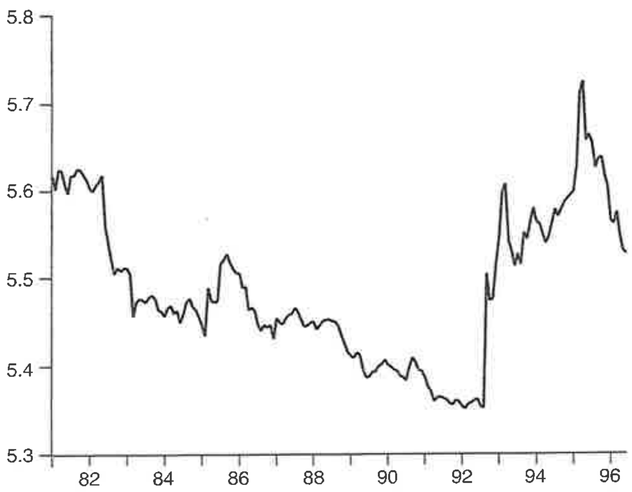

Figure 8.6 Log real exchange rate Italy–France, 1981:1–1996:6

Table 8.4 Unit root tests for log real exchange rate Italy–France

Statistic	Without trend	With trend
DF	−1.930	−1.942
ADF(1)	−1.874	−1.892
ADF(2)	−1.930	−1.961
ADF(3)	−1.987	−2.022
ADF(4)	−1.942	−1.981
ADF(5)	−1.966	−2.005
ADF(6)	−2.287	−2.326

8.6 Estimation of ARMA Models

Suppose that we know that the data series $Y_1, Y_2, \ldots, Y_T$ is generated by an ARMA process of order p, q. Depending upon the specification of the model, and the distributional assumptions we are willing to make, we can estimate the unknown parameters by ordinary or nonlinear least squares, or by maximum likelihood.

8.6.1 *Least Squares*

The least squares approach chooses the model parameters such that the residual sum of squares is minimal. This is particularly easy for models in autoregressive form. Consider the $AR(p)$ model

$$Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + \varepsilon_t, \qquad (8.61)$$

where ε_t is a white noise error term that is uncorrelated with anything dated $t-1$ or before. Consequently, we have that

$$E\{Y_{t-j}\varepsilon_t\} = 0 \quad \text{for } j = 1, 2, 3, \ldots, p,$$

that is, error terms and explanatory variables are contemporaneously uncorrelated and OLS applied to (8.61) provides consistent estimators. Estimation of an autoregressive model is thus no different than that of a linear regression model with a lagged dependent variable.

For moving average models, estimation is somewhat more complicated. Suppose that we have an $MA(1)$ model

$$Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1}.$$

Because ε_{t-1} is not observed, we cannot apply regression techniques here. In theory, ordinary least squares would minimize

$$S(\alpha, \mu) = \sum_{t=2}^{T} (Y_t - \mu - \alpha\varepsilon_{t-1})^2.$$